DOCUMENT RESUME

ED 214 967 TM 820 144

AUTHOR	Duckett, Willard R., Ed.
TITLE	Planning for the Evaluation of Teaching. NSPER: 79. A CEDR [Center on Evaluation, Development and Research] Monograph.
INSTITUTION	Phi Delta Kappa, Bloomington, Ind. Center on Evaluation and Research.
PUB DATE	[80]
NOTE	101p.; Small print throughout.
AVAILABLE FROM	Phi Delta Kappa, P.O. Box 789, Bloomington, IN 47402 ($5.00).
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Data Collection; *Evaluation Methods; Evaluators; Teacher Evaluation; Teachers
IDENTIFIERS	Evaluation Problems; *National Symposia for Professionals Evalu Research

ABSTRACT

This is a series of papers delivered at three National Symposia for Professionals in Evaluation and Research (NSPER) sessions in 1979. The agenda was the same at all sessions. The main topic was "Planning for the Evaluation of Teaching." The sessions were conducted in Charlotte, North Carolina; Milwaukee, Wisconsin; and Albuguerque, New Mexico. The participants agreed upon the importance of planning, the situation-specific nature of evaluation, and its dynamism. They believed a statement of philosophy concerning teacher evaluation is important. An examination of evaluation instruments for teacher evaluation indicated an emphasis on organizational maintenance rather than teaching improvement. A responsible evaluation must include formative, process, and summative considerations. The primary sources of data to be used include students, colleagues, administrators, and teachers' self reports. The secondary sources for data collection include observation, interviews, and rating scales. Feedback procedures such as audiotape, videotape, minicourses, interaction analyses, and student description are helpful in improving instruction. The editors and co-directors of NSPER: 79 developed a summary of the sessions, "The Basis for Evaluating Teaching." (DWH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED214967

PLANNING FOR THE EVALUATION OF TEACHING

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D. Kliewer

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

TM 820 149

2

Planning for the Evaluation of Teaching

Edited by

WILLARD R. DUCKETT
Phi Delta Kappa

Contributing Editors and Co-Directors
of NSPER:79

WILLIAM J. GEPHART
Phi Delta Kappa

ROBERT B. INGLE
University of Wisconsin-Milwaukee

MARY R. CARROLL
Phi Delta Kappa

Cover Design: Mary Seldon

Table of Contents

PREFACE v

NSPER:79 — PLANNING FOR THE EVALUATION OF TEACHING vii

NSPER:79 — CHARLOTTE

Planning for the Evaluation of Teaching
James B. Macdonald 2

Evaluation of Teaching: Purpose, Context, and Problems
James B. Macdonald 13

NSPER:79 — MILWAUKEE

Planning for the Evaluation of Teaching: Guidelines and Principles
Dale L. Bolton 28

The Bases for Evaluating Teaching: Philosophy, Context, and Purpose
Dale L. Bolton 39

NSPER:79 — ALBUQUERQUE

A Different Approach to Teacher Evaluation
Michael Scriven 60

Teacher Evaluation: The Myth and Realities
Carolyn J. Wood and Paul A. Pohland 73

SUMMARY

Wrap-Up: The Basis for Evaluating Teaching
William J. Gephart, Robert B. Ingle, and Mary R. Carroll 84

Preface
NSPER: 79

Every time we teach we are evaluated. We have no choice in that. That evaluation is used by students in decisions they make. We have no choice in that either. And further, that evaluation by our students is sometimes passed on to others — parents, other teachers, administrators — and is used in decisions they make about us. We have no choice in that either.

Our choices about having our teaching efforts evaluated lie in the manner that evaluation is done. We can do nothing and the private, subjective, accidental, and often incomplete evaluation cited above will continue. Or we can assert ourselves as professionals and structure public, objective, systematic, and comprehensive evaluations of our teaching. If we do that the private, subjective evaluations will continue. But, decision makers will have additional information, information we can more often support, to use in decisions about our teaching.

If public, objective, systematic, and comprehensive evaluation of teaching is to be of maximum value, two things need to be considered—the bases on which the evaluation rests *and* the thoroughness and adequacy of planning for the evaluation and its use. What is the philosophical base on which the evaluation rests? Are people important? Are numbers important? Accomplishments? What is the context in which the evaluation of teaching will take place? What purposes can be (will be) served through the evaluation of teaching? These and related questions are crucial. They help us determine how we will participate in the evaluation of our teaching.

Planning the evaluation of teaching is where we can determine whether evaluation is a tool that is potentially punitive to us or helpful and whether it is useful in the decisions others make about us, decisions we make, or both. Evaluation involves a variety of people doing a variety of things with a variety of equipment. If it is to be maximally useful it must be planned carefully and thoroughly.

NSPER: 79 focuses on those two points—the basis (philosophy and purposes for evaluation) and the planning for the evaluation of teaching.

Clearly, how we will evaluate teaching is a problematic question, BUT we must do it. NSPER: 79's focal objective is to involve educators in the development of a statement that shows

how the philosophy, contexts, and purposes for the evaluation of teaching might be examined, and how that information can be used in planning.

Planning for The Evaluation of Teaching

Since 1974 a number of activities of Phi Delta Kappa's Center on Evaluation, Development and Research have been oriented toward the evaluation of teaching. The National Symposia for Professionals in Evaluation and Research (NSPER) that year included presentations on the nature of teaching and involved the participants in a Teacher Appraisal for Improvement (TAFI) Workshop. Developed by the UCLA Center for the Study of Evaluation, the TAFI Workshop acquaints participants with eight different approaches for measuring the quality of teaching. NSPER: 75 focused on the evaluation of administrative performance. In late 1976 Phi Delta Kappa contracted with UCLA to organize and conduct TAFI workshops. As of July 1980, 92 of these workshops have been conducted in various locations in the United States and Canada. Twenty-one more are scheduled during the 1980-81 academic year. NSPER: 78 focused on the assessment of student competence in the public schools, and as such was a focus on the results of teaching.

All of these activities contributed evidence to the fact that evaluation of teaching happens—all the time. Much of the time that evaluation of teaching "just happens," it's product—information about the quality of teaching—is subjective and of questionable reliability; not what most of us would like to see used in decisions about our future as a teacher. Some of the time evaluation happens as a deliberate process. If done well on those occasions, it produces information that can help us in our decisions about how we teach, and it can help administrators and supervisors with the staff personnel decisions for which they are responsible. Unfortunately, the occasions on which the evaluation of teaching is done well are infrequent. Or at least, that is the predominant perception. Few teachers can be found who willingly acknowledge that they were evaluated in a way that helped them improve their teaching—very few.

The deliberate planning for evaluation of teaching seems to be the crucial factor. Good evaluation does not "just happen." It is carefully planned. But, what are the components and processes central to that planning? If good planning is the ingredient that can make evaluation of teaching a useful, humane activity, how can we plan? Are there specific things we should attend to? Are there particularly useful structures or procedures

for planning for the evaluation of teaching? These questions and others like them became the focus of the NSPER: 79 sessions, "Planning for the Evaluation of Teaching." As in the 12 previous NSPER sessions, two objectives structured the planning for the 1979 sessions in Milwaukee, Albuquerque, and Charlotte. Number one was to help educators responsible for the evaluation of teaching understand the importance of, and the basis for, planning for the evaluation of teaching. Number two was to provide a vehicle for individuals to share information on their successes and failures in performing their evaluative duties. The provision of numerous opportunities for small group discussion and opportunity for unstructured interactions in the evenings was the basis for achieving objective number two. Procedures for achieving number two were a bit more complex and are described in the next section of this report.

The three NSPER: 79 sessions, as has been the tradition, had the same agenda but a different cast of characters in the participant and major resource roles. They took on distinctly different character and yet were much the same. The three sessions contributed "pieces" to a structure that seems logically useful in planning for the evaluation of teaching. A mnemonic for that structure is the numbers one, two, and three. There is *one goal* that justifies the expenditure of resources in the evaluation of teaching, the improvement of student growth (in cognitive, affective, moral, and psycho-motor dimensions). There are *two* evaluative approaches that are used to contribute to that goal. Sometimes we evaluate my teaching to give *me* information *I* can use in *my* decisions about how I am going to teach—evaluation for individual professional development. Sometimes we evaluate my teaching to give *administrators and supervisors* information *they* can use in *their* decisions about how they are going to use me—evaluation for personnel management. There are *three* classes of variables that are or should be used in describing the teaching that is to be evaluated: teacher style variables (What is the approach, techniques, procedures used by the teacher?), employee variables (punctuality, dedication, ability to work effectively with others, etc.), and context variables (items established by the location in which teaching takes place—e.g., the nature of the learners, state laws and local policies that regulate what can or cannot be taught, policies regulating length of instruction, etc.).

Planning for the evaluation of teaching has been and will continue to be a concern for educators and education's publics. Efforts to improve our planning in this area will not and should

not eliminate that concern. We will never be finished with that planning task. Nor will we find the adoption of someone else's plans an easy way out. Despite those points it can be a rewarding area of endeavor.

WJG
RBI
MRC

NSPER: 79

Charlotte Presentation

JAMES B. MACDONALD
University of North Carolina
at Greensboro

Planning for the Evaluation of Teaching

James B. Macdonald

It should be clear from my review discussion above that the evaluation of teaching is a complex and difficult task. But, my previous remarks were spoken as a theorizer and researcher in the field. From those perspectives I can only say that we do not really know enough to do an adequate job of evaluating teaching, especially in relation to life determining decisions such as merit, promotion, and tenure.

But like all educators, I suffer from the professional schizophrenia of being a practitioner as well as a researcher and theorizer. As such I respond to political and social pressures from outside the profession that place practical problems in our laps. The reality of today's practice leads us to the evaluation of teaching whether we wish it, or know it to be desirable, or not. We are in an era of what is called "accountability." Literally speaking we are now being asked to account for public expenditures on education in terms of acceptable outcomes.

It is clear that the total cost of our educational endeavor in America has increased rapidly over the past few years. It is also clear that the increased costs cannot be traced solely to teacher salaries. While costs have risen dramatically, the public's attitude toward education has been changing. At least three major perspectives have developed.

1. *The "re" discovery of socio-economic relationships to achievement.* It really was not until the mid-Sixties and on into the 1970s and up to the present day that the general public became aware of the fact that large numbers of lower class and lower middle class youngsters were not doing well in schools.

This has been called a "failure" of our schools, and perhaps there is some truth to this allegation. But the truth of the matter is that historically and cross culturally these kinds of facts have been known in the profession for decades. The problem was the same in 1850 and is the same in Sweden, France, England, etc., as in the United States.

Thus, the discovery that large numbers of students do not do well in school (45% of New York City students and approximately 25% nationwide drop out of high school) is causing a re-evaluation of the value of education, and a call for "accountability."

2. *The achieving student is doing less well today.* We probably could survive the rediscovery discussed above if there weren't also evidence that students who formerly performed at satisfactory levels are scoring lower on standardized tests than they did 10-15 years ago. Data on decreasing SAT scores, for example, are now well known among elements of the community that have direct access to decision making through boards of education and powerful community pressure groups. Thus, the call for accountability is not just from groups whose children historically do poorly in school, but from groups whose children traditionally do well in school.

3. *The taxpayer's revolt.* Put these together with the taxpayers' revolt which, of course, directly affects schools and you have a generalized citizens' revolt against increasing costs focused on one of the few agencies in society that peoples' wishes can affect directly—the educational system.

This pressure is felt differently in public schools than by colleges and universities, but the pressure is there and judgments affecting the lives of large numbers of teachers are being made constantly at both levels. These public political pressures on the schools exist even though:

1. home backgrounds may contribute far more to achievement than school, and
2. the school has *multiple functions*, such as
   a. socialization,
   b. indoctrination,
   c. caretaking (baby-sitting), and
   d. keeping youth out of the labor market.

In fact some critics suggest school is working just as it should, producing *winners* and *losers*, replicating the status quo of the class system while producing a meritocracy that society needs.

These factors mean that we educators are faced with an awesome responsibility. We must, knowing that we do not know a great deal about evaluating teaching, make our evaluations in the fairest and most defensible manner we can. Under these circumstances, common sense tells us that any appraisal system that we implement must be comprehensive enough to evaluate multiple criteria (e.g., formative, processing product, summative tests, teachers content knowledge and teaching skills) using multiple methods (i.e., systematic observation, peer supervision, student ratings, self appraisals, and indices of pupil achievement).

Multiple Criteria

Let us look at the kinds of data we must collect to perform a reasonable evaluation of teaching.

Teachers make many proactive or formative decisions. This varies, of course, with the level and focus of the teaching. But the act of planning is necessary at all levels. So, areas that require planning by teachers become sources of *formative data*. As such, plans, syllabi, selection of supplemental materials, range and variety of alternative activities and samples of testing forms and procedures (tests and plans for evaluation) are all formative data that may be examined.

Teacher-pupil interactions provide a second kind of data that form a basis for making judgments about teaching. Thus, data concerning the way in which teachers interact with pupils is an important aspect of evaluating teaching. Our concern here is for the achievement of students in relation to the goals of instruction. These goals, it should be remembered, are also *multiple goals*, including cognition, attitude and affect, as well as caretaking.

The point that must be remembered is that a responsible evaluation of teaching will, by necessity, involve formative, process, and summative considerations.

Multiple Sources of Data

Not only are multiple criteria necessary, but perspectives on teaching must come from multiple sources. Our current knowledge suggests that there is no single source or procedure or single critical criterion that can be depended on by itself. Thus, we are forced by responsible evaluation concerns to utilize a variety of sources of data in evaluations. There are four primary sources and a number of secondary sources that are important to consider. The primary sources provide minimum data for evaluation, while the secondary sources provide a full range of meaningful and adequate data.

Primary Sources

1. *Students.* It is clear that students see more of the teachers than anyone else. They live with the teacher and experience the personal and professional aspects of the teacher on a day-to-day

basis. Thus, students are primary sources of data. Students at all ages can participate in some meaningful way in supplying important data about teachers.

Student rating scales can normally be used by the fourth grade level, continuing through the college level. The rating scales can be constructed specifically for each level, discipline, or unit in the system, or they can be standardized for all units or levels with modifications. With younger children interviews or reaction sheets can be of value.

Research data indicate that student ratings of teachers are subject to some "popularity" pull or "entertainment" pull, but still correlate reasonably with achievement. What data we have tells us that student opinions and judgments do partially reflect the amount of learning that takes place. Students obviously are in no position to evaluate the formative or proactive aspects, i.e., planning and preparation for teaching, but are probably best able to evaluate interactive aspects of teaching which is, at the very least, a valuable perspective on teaching.

2. *Teaching Peers.* Colleagues and peers are seen as a valuable source of evaluation data, provided a systematic approach (rather than teachers' room gossip) is utilized in the process. Peer evaluations of formative data are especially useful. Our colleagues can make fairly clear judgments about the kinds of plans and proactive decisions we make as teachers.

Some writers, in discussing techniques of teacher evaluation, have noted, however, that a lack of operational definitions, failure to control for sampling of teacher behavior, and the effect of the observer on teacher performance can create limitations that make rating scales of doubtful worth in the hands of administrators, supervisors, and *peers*. Thus, as a primary (i.e., readily available) source of evaluation, some doubt is cast on peers and administrators.

3. *Administrators and Supervisors.* Administrators are primary sources of evaluation because the role of evaluator is allotted to administrators most frequently in the schools. This has come about as school practice has been patterned more and more after business practice, with the creation of employer-employee relationships between administrators and teachers.

The advantage of this arrangement is that administrators continuously receive informal data about teachers. Disadvantages include the possibility that this data will be filed unexamined and unreflected upon, or not presented to the specific teacher from whom a reaction and comment seem warranted.

Administrators are poor evaluators for several reasons, primarily because they are in a hierarchical power relationship to the teacher. This relationship can distort the teacher's behavior in the presence of the administrator resulting in an inadequate evaluative session. Quite truly, we would be better served if we trained objective observers, with *no* institutional axes to grind, to collect data through observation.

4. *Self reports.* In self reports we essentially ask teachers to supply us with descriptions of their teaching. They may be in the form of pros and cons or strengths and weaknesses, a listing of problems, or check lists, or what have you. Under any circumstances, self reports have limited use. A major weakness is that most teachers present descriptions that look good on paper. Further, teachers *by themselves* have not been shown to have any special awareness of difficulties in their own teaching. Studies show that teachers think they are doing fine until they are presented with direct evidence that they are not meeting standards that they themselves accept. Then, most teachers are open to "seeing" and changing!

Thus, we see that the primary sources are more or less useful. Students seem to be more helpful than peers or administrators.

Secondary Sources

Secondary sources for collecting data in the evaluation of teaching are less directly available and usually call for extra expense and larger blocks of time. They may, however, be better sources of data. For example:

1. *Systematic observations.* These should be made by trained observers.

2. *Contract plans.* These should be assessed by looking at student gains; or single lessons observed by peers with a single educational or teaching goal in mind.

3. *Performance tests.* These are tests of teaching power. The teacher is given one or more objectives (the same one or ones for all teachers) and each teacher devises his/her own instructional tactics. A group of learners is randomly assigned to the teacher, usually a small number, less than the typical class size, and tests are given at the end of instruction. Comparisons are made of post-tests across teacher means.

4. *Domain-referenced testing.* These are used for testing student achievement using multiple matrix sampling.

5. *Observer-written protocols describing teachers' work.* These protocols are examined by trained raters who rate them using appropriate criteria.

Feedback Procedures

Allied to the secondary sources discussed are a number of feedback procedures helpful in improving instruction. The idea here is that teachers need "mirrors of conduct" to reflect their own behavior back to them. From this they can single out aspects about which they may decide to develop a plan for improvement.

1. *Audiotape.* Teachers or an observer may tape record a session of a given length, or a number of sessions in a variety of fields. After the lesson teachers, either alone or with an observer, listen and discuss the tape(s). From this, some plan of improvement is developed and the teacher sets up a taping at some later date to check on improvements.

2. *Videotape.* Videotape can be used much as audiotape, or it can be used in such procedures as micro-teaching. In micro-teaching a teacher is taped for 5-10 minutes with a selected group of youngsters while demonstrating or working on one skill at a time. As feedback, videotape gives teachers considerable opportunity to learn.

3. *Minicourses.* Minicourses are packages of materials, including manuals, self-administered tests, and films intended for use with videotape or audiotape recorders furnished by the school. With a typical minicourse the teacher can acquire training in a particular skill such as designing independent learning experiences for primary grade children (Borg, 1970). First, the teacher reads about skills involved in the handbooks. Then he/she sees a film explaining and illustrating the skills. Next, a lesson is planned by the teacher for practice. Finally, the teacher evaluates the lesson.

4. *Interaction analyses.* There are a great number of different systems of observation available for use. The best known (though not necessarily the 'best') such system was developed by Flanders (1970). Here, a trained observer (an administrator, supervisor, or colleague) observes a teacher and, using a code, tallies the behaviors present. Afterwards a summary is made and the teacher's behavior is discussed by the observer and the teacher. On the basis of this discussion, goals for future improvement are generated.

5. *Student description.* Student descriptions of their teachers can be used as feedback to teachers. They are typically generated through the use of rating scales. These scales provide another "mirror" for teachers to use in improving their teaching.

Summary

Sources of data that may be used in the evaluation of teachers include students, colleagues, administrators, the community, the alumni, and self reports. Techniques that may be used to collect this data are observation, (formal and informal tests, surveys, opinionnaires, etc.), interviews (formal and informal), rating scales, and self reports. Naturally, in any measurement activity we must be concerned about the validity, reliability, objectivity, and practicality of any technique. There are difficulties for one or more of these criteria in all of the approaches discussed. As a rule of thumb, one can say that *no procedures developed up to this time adequately meet all four criteria.*

Impediments to Measuring Teacher Effectiveness

Before leaving the measurement topic, I would like to discuss certain impediments that exist in the process. David Berliner's article (1976) is helpful in this respect. Berliner thinks persons who are pushing for teacher evaluation in accountability systems suffer from "ostrachism," a common disease affecting education. Its definition, or etiology, is a premature commitment to a particular educational movement. One symptom of this is sticking one's head in the sand when problems appear, in the hope that the problems will go away. Essentially what Berliner describes as current "ostrachism" is the linking of teacher competence to student outcomes and the lack of convincing empirical evidence for this linkage.

As mentioned before, we lack conclusive evidence that teachers are the substantial causative effect on student achievement, yet the heart of the accountability argument rests on such linkage. The purpose of improvement of instruction can proceed without this assurance. There are serious problems with this assumed linkage. For example, there are major instrumentation problems related to the outcomes, the dependent variables of schooling. Serious problems exist with standard-

ized testing, special unit tests, and in developing multivariate outcome measures. Standardized tests, for example, may not accurately reflect what is taught in any one teacher's room, and thus are not a valid measure of student outcomes. They are also highly correlated with intelligence tests. And, their multiple choice format may not be valid for minorities, the young, and bilingual children.

The measurement of "process" (independent) variables also causes grave difficulties. The usual approach is to identify some positive behavior like "asking higher order questions," or giving verbal praise, etc., and then counting the number of times that behavior appears. However, when considered out of context of the total situation, a specific behavior makes little sense. It is the quality of the merged use of certain behaviors in light of the total situation that makes sense. This, as yet, cannot be coded and qualified.

Another problem is the so-called "unit of analysis." Is the unit in teaching process measurement the teacher question? Or, is it the teacher questions *and* wait time? Or, teacher questions, wait time, *and* student answers? If the latter, are student questions part of an even larger strategy that should be the unit of analysis? One thing seems certain, the idea that you can isolate and observe a behavior does not establish that behavior as a good unit for analysis. It may mean nothing by itself!

A third problem with independent variables is the stability of teaching behavior. If we are to observe, we must know that a behavior will occur sometime during observation and that this is the teacher's customary or usual way of behaving. We must also recognize that the same behavior will probably not appear at the same time when that lesson is taught again by that system, the results will show very low correlations in stability (one study -.90) over two times. Thus, the "indirect" or "direct" teaching seems highly unstable. If that is the case, what do we deem the teachers "usual" behavior to be? Some behavior seems more stable than others. But, we are beginning to realize that many important teaching behaviors seem unstable.

There are also *methodological* problems beyond those of instrumentation.

1. Student background is highly correlated with achievement. Thus, achievement alone is difficult to use as an outcome or product variable.
2. Different subject areas are more or less related to teacher effectiveness. Achievement in reading and social studies

are less related to teaching than achievement in science or mathematics.

3. Research using norm-referenced standards is typically done with volunteer teachers and thus is skewed. When judged against each other, some teachers look better; however, all may be terrible or great. One cannot tell in most research.
4. Teachers adjust to individual differences, but our observation or research looks at groups. In doing so, it makes no allowances for individual differences among students.
5. Student time on task is a critical variable, but the teacher doesn't influence this directly. We don't know which behavior produces "time on task."
6. Construct validation and effectiveness must mean the same to everyone. This is rarely true in research.
7. The generalizability of effectiveness measures seems to be reasonably stable when the same teacher teaches the same material to the same kinds of students. But change the material or the students and effectiveness measures are not stable. Thus, *effectiveness is context bound;* teachers are not generally good, bad, or indifferent but are specifically good, bad, or indifferent. Effectiveness isn't like a birth defect; it's like sunburn (specific to context).

The data from studies in elementary grades in relation to process-product relationships support what Rosenshine (1975) calls "direct instruction."

Critical aspects of "direct instruction" are that the teacher

1. focuses on academic goals,
2. promotes extensive content coverage and a high level of pupil involvement,
3. selects instructional goals and materials and actively monitors student progress,
4. structures learning activities and includes immediate, academically oriented feedback, and
5. creates an environment that is task oriented but relaxed.

According to Brophy (1979) several studies provided strong support for the following generalizations:

1. Teachers make a difference.
2. Even so, there is no support for the existence of generic teaching skills, very specific behaviors. There is some evidence in support of higher level relationships.
3. One such cluster includes teaching expectations and role definitions.

4. Another basic pattern includes such variables as classroom management, student engagement, time on task, and student opportunity to learn. These are associated with increased instructional time.
5. There is a cluster of support for direct instruction lectures, recitation, drill, practice, and the teacher working with the whole class.
6. The strength of support for these clusters varies with context, grade level, and teacher ability level.

Thus, for low-level cognitive learning, it looks like a traditional model for teaching is the best.

A recent study by Coker, Medley, and Soar (1980), however, raises questions about Rosenshine and Brophy's conclusions. In a fairly extensive look at 100 classrooms they attempted to find relationships between teacher competencies and pupil gains in achievement and self concept. They summarized as follows:

> The inconsistency of relations with different outcomes; the frequent violations of expectations, sometimes only at certain grade levels; and the number of competencies that failed to relate at all—these lend little support to our common assumption that beliefs of teachers or experts about the nature of effective teaching are generally correct. (p. 149)

Where this leaves us, I believe, is with a great humility about making judgments about teacher effectiveness from the kinds of knowledge and skill we presently possess. This is why I said earlier that common sense tells us to use multiple techniques and multiple sources. Then we can look for patterns and maybe some will emerge.

Michael Scriven (1977), on the other hand, seems more confident than Berliner or I. He proposes the T-form with some verbal optimism. He says there are several legitimate inputs. They are:

1. Student estimates of teaching merit
2. Learning gains due to the teacher
3. Subject matter choice and coverage, treatment testing, and feedback
4. Estimates of classroom justice or injustice, e.g., fairness, sexism, social classism, due process, etc.
5. Non-classroom contributions to teaching

These may be obtained as follows:

1. Student form with a single question, calling for the evaluation of a single skill (at a time) on an A-F scale
2. Pre-post tests

3. Subject matter specialists, colleagues
4. Examine S Forms, peers, specialists
5. Administrators plus peers, possibly students

Conclusions

In closing I must say that my theory and research orientation tell me to go slowly. I would be pleased with a moratorium on teacher evaluation—as with IQ and standard achievement tests.

Yet, someone will have to do it. If you can't avoid it, the only possible thing to do justifiably is to use multiple methods and sources and pray that you can put it all together.

What I personally believe is possible with our present knowledge is the identification of a small group of inadequate teachers and a small group of excellent teachers. I do not believe we are capable of making distinctions among those in the large middle group.

References

Berliner, David. "Impediments to the Study of Teacher Effectiveness," *Journal of Teacher Education*, Spring 1976, pp. 5-13.

Borg, Walter R. *The Mini Course: A Micro-Teaching Approach to Teacher Education*. Beverly Hills, Calif.: MacMillan Educational Service, 1970.

Brophy, Jere E. "Teacher Behavior and Student Learning," *Educational Leadership*, October 1979, pp. 33-38.

Coker, Homer; Medley, Donald M.; and Soar, Robert. "How Valid are Expert Opinions About Effective Teaching?" *Phi Delta Kappan*, October 1980, pp. 131-134, 149.

Flanders, Ned. *Analyzing Teacher Behavior*. Reading, Mass.: Addison-Wesley, 1970.

Rosenshine, Barak. "Classroom Instruction," in N. L. Gage, ed., NSSE 75th Yearbook, *Psychology of Teaching Methods*. Chicago: University of Chicago Press, 1975.

Scriven, Michael. "The Evaluation of Teachers and Teaching," in Gary D. Borich, ed., *The Appraisal of Teaching*. Reading, Mass.: Addison-Wesley, 1977.

Evaluation of Teaching: Purpose, Context, and Problems

James B. Macdonald

The evaluation of teaching is a professionally "active" topic. When considering a topic as active, it is probably well to remember some previous active topics. Where are they now? Most of these previous movements (like open education and multi-age groupings) were very active for a time and were then integrated in some meaningful way into the fabric of schooling. So the present intensity of the pressure to evaluate teaching will subside eventually. Meanwhile, we should take advantage of the present opportunity. While the topic enjoys political vitality, we need to locate the critical professional concern in the evaluation of teaching movement so that we may integrate that part into our practices when the "heat" subsides.

Practically every thinker and writer on the topic begins the discussion with purposes of teacher evaluation. Further, there is near unanimous agreement that the purposes are two-fold:

1. evaluation of teaching to facilitate the improvement of instruction, and
2. evaluation of teaching to facilitate institutional/administrative decision making (whom to promote, whether to grant tenure, whether to give salary increases based on merit, etc.).

Certain difficulties immediately arise. Logically, one could argue that the two purposes are not mutually exclusive; that the same data can be used for both purposes. And, logically, this appears to be valid. But the psychology and sociology of the teacher make this logic look rather dubious. It soon becomes clear that, from the teacher's perspective, the context of each purpose is quite different. Thus experts tell us that the two purposes should never be mixed, and that teachers should be aware at all times of the purposes for which data are being collected.

Let me quickly illustrate the problem. Assume that an administrator has visited a classroom to make observations and has used a well-known observation scale—OSCAR, for example. If the purpose is for administrative judgment, then the validity, reliability, and objectivity of the scale and data collection procedures are of paramount importance. That is, a person does

not wish to be judged by some techniques or processes that cannot be demonstrated to be good ones. After all, life-affecting decisions may be dependent upon it.

On the other hand, if the data were to be used for improving instruction, the context is quite different. More than likely the results of the observation would be shared with the teacher (no judgment *per se* is necessary, just sharing data). The teacher is not threatened because he or she will take into consideration these results, this feedback on teaching, and build this new knowledge into further improvement. Freedom to reject the data are present in the process, too. Thus the data either help or they do not, depending on the teacher and other factors. If they do not help, other procedures will be used. At no point will potentially threatening job survival decisions be made by an outside agent.

It seems clear that most professionals believe that the more significant and important purpose of evaluating teaching is for the improvement of instruction. In the practical context and institutional setting, however, the data are more apt to be used for administrative decisions.

Let us look at what we mean by the evaluation of teaching. In general terms, it refers to a process we use in making judgments about someone's teaching. There are at least three steps involved. First, it is necessary to experience in some fashion the teaching, either your own or that of another; next, it is necessary to interpret and understand the meaning of the experience in its context, and finally to make some judgment about it.

Experiencing. Evaluating teaching means that we (the evaluators) experience the teaching in some manner, usually and typically in a variety of ways. Thus, administrators may ask teachers to submit plans, samples of tests, etc. They may give tests which tap the knowledge of the teacher and students on a number of points; they may observe the teacher in the process of teaching; and they may hear comments from students, parents, other teachers, etc. All of these avenues and many other possibilities allow the evaluator to experience teaching of the teacher being evaluated.

Interpreting. Experiencing is only the first step. One must interpret what one has experienced, i.e., put it into perspective and context in order to understand its meaning for the evaluation of teaching.

Interpreting experience may begin before the actual teaching occurs in that the selection of techniques for collecting data will

specify significant values and areas to watch for during the evaluation period. After the evaluation experience, the data are categorized in many cases, i.e. interpreted for meaning. However, other kinds of interpretation must still go on. The meaning of the experience must be examined in relation to ethical, aesthetic and other holistic phenomena and criteria. This takes place during and after experiencing.

Judging. The third step in the process involves judgment. It is at this point of the process where values become most prominent. Judgment demands holding up significant values or standards and comparing performance in relation to these values. It is a complex process not well understood. It concentrates on putting together data of specific and global kinds, experiences in a variety of ways, over a considerable period of time. This must usually (for administrative purposes) be translated into a fairly singular quality or quantity for use. For the purpose of improving instruction, the data may simply be left available and judgment then becomes the teacher's task (perhaps with the administrator), not primarily the administrator's judgment.

Organizational Philosophy in Operation

In many ways the evaluation of teaching is affected by the philosophy in operation in the organization. Two predominant organizational models exist—the mechanistic model and the humanistic model.

Mechanistic model. The mechanistic model basically accepts the idea that teaching is a collection of skills, knowledge, and attitudes that have been learned and held in the repertoire of the teacher. There is, thus, some critical level of possession of competence that allows us to say, "This teaching is competent, good" or, "This teaching is incompetent, bad" or somewhere in between. In the mechanistic organizational model the evaluation of teaching is essentially concerned with the assessment of the "parts" of teaching. In essence, the focus is primarily that of identifying poor or faulty parts that must be replaced, restored, or fixed in some manner; corrections that will lead to more competent functioning.

In the mechanistic organization teachers are essentially objects of evaluation, and great emphasis is placed on the objectivity of assessment procedures. The basic interest is to remove all bias or intuitive judgment from the proceedings. The focus is on the objective reality of the objects (teachers and pupils) in

interaction. The objective data are collected and used to make decisions about assignment, promotion, and tenure.

Humanistic model. The humanistic model begins with a central commitment to the people involved. It deliberately tries to avoid treating teachers as objects. Each person is seen as an individual with dignity and integrity, as a sharer of the human condition or mortality. Each teacher is perceived as a person in the process of becoming, of creating meaning in his or her life. As a faculty, they are persons striving to do the best they can for themselves and others in their lives and profession. As such, teachers and administrators are colleagues in the humanistic organizational model; collaborators in a profession with shared goals and common purposes and hopes for both their students and themselves.

The evaluation of teaching in the humanistic organization is seen as a way of providing feedback or information to teachers to help them as persons engaged in improving their instructional efficiency. When an individual seems to be misplaced in the profession, there is an effort to provide value clarification and counseling that will assist in the consideration of a profession other than teaching.

The evaluation of teaching in the humanistic organization is much less a concern for quantity of competence and much more concern for the quality of the total classroom atmosphere and environment. The emphasis is as much or more on the *quality of living* in classrooms as it is with the *quantity of achievement.*

Is Teaching An Art Or A Science?

A central concern is whether or not teaching is an art or a science. Can teaching, for example, be analyzed and discussed in terms of the laws or principles of a science or is it an expressive art that does not lend itself to analysis?

Nate Gage (1977) approaches this problem by stating that what we need is the discovery of basic scientific principles in the art of teaching. He, in other words, does not believe that there can be a science of teaching *per se*, but that certain basic scientific principles are necessary and will be useful for the art of teaching. Gage says:

> Let me begin by distinguishing between a science of teaching, and a scientific basis for the art of teaching. The former idea, a science of teaching, claims much more and is in the end, I think, erroneous. It implies good teaching

> will some day be attainable by closely following rigorous laws that yield high predictability and control. (p. 17)

Gage goes on to say: "Practical enterprises, those conducted in the real world rather than the laboratory, have both artistic and scientific components." (p. 17)

In a sense, I believe Gage is saying that practical enterprises, such as teaching, are activities that have social, ethical, and political elements as well as technical skills. Practical does not mean merely technical. It means ethical and personal efforts of the individual teaching, calling for intuition, creativity, improvisation, and expressiveness.

At this point, Gage opts for a "scientific basis for the art of teaching." This is his interest. He says, in passing, that the artistic elements of teaching might be assessed with techniques similar to those suggested by Elliot Eisner (1978), akin to the procedures used by art critics. Connoisseurship is the term Eisner uses. (p. 133)

In the study of holistic qualities, critics and connoisseurs use the ability to perceive integrated patterns relying on their accumulated experience, cultural background, and evocative language to grapple with the task of evaluation. What this material suggests is that we should look for process-product variables and relationships to find teacher behavior related to pupil achievements.

Range of Variables In Teaching

For Gage, it is useful to think in terms of four kinds of variables that must be accounted for: presage, context, process, and product variables.

Presage. These variables are characteristics of teachers, such as age, sex, socio-economic background, personality, and educational experience. Teachers differ on these variables. Therefore, knowledge of relevant teacher characteristics is useful when evaluating teaching. For example, the socio-economic background of the teacher is useful knowledge when evaluating a teacher's interactions with specific socio-economic groups of youngsters. And, of course, personality, educational background, and experience are obviously related.

Context. Context variables are similar to Lundgren's (1979) "frames." Conditions such as objectives, organizational policies, and student characteristics "surround" teaching and provide a context for teacher behavior. Context variables have important effects on teaching.

Process. These variables cover the interaction behaviors of teachers and students. Teacher performances, competence, skills, behavior, or actions are all perceived as process variables. What teachers do in the classroom is clearly seen as important in the evaluation of teaching.

Product. These variables include the learning outcomes of students, i.e., student achievement, in classroom situations.

The critical variables are *process* and *product*, how teachers behave and how this relates to pupil outcomes. The scientific basis consists of looking for non-chance relationships in the realm of events with which practice is concerned. Relationships need not be perfect or even close to perfect. Rather, relationships need merely to be better than those that would occur by chance.

Reviewing the research on teaching, Gage believes that we may be able to come up with teacher "shoulds" based on the study of process-product relationships. His examples of relationships relevant to studies of *primary school programs* (which he says are consistent with conclusions by Rosenshine (1971), Medley (1959), and others) are:

1. Teachers should have a system of rules that allows pupils to attend to their personal and procedural needs without having to constantly check with the teacher.
2. Teachers should move around the room a lot, monitoring pupils' seatwork *and* communicating to their pupils an awareness of their behavior, while also attending to their academic needs.
3. When pupils work independently, teachers should insure that the assignments are interesting and worthwhile yet still easy enough to be completed by each learner working without teacher direction.
4. Teachers should keep to a minimum such activities as giving directions and organizing the class for instruction. Teachers can do this by writing the schedule on the board, insuring pupils' knowledge of where to go and what to do.
5. In selecting pupils to respond to questions, teachers should call on a child by name *before* asking the question and assure that all pupils are given an equal number of opportunities to answer questions.
6. With less academically oriented pupils, teachers should always aim at getting the child to give some kind of response to the question. Rephrasing, giving cues, or asking a new question can be a useful technique for bringing

forth answers from a previously silent pupil or one who says, "I don't know" or answers incorrectly.

7. During reading group instruction, teachers should give a maximal amount of brief feedback and provide fast-paced, "drill" type activities.

Essentially what these findings lead to is provision for maximum academic learning time. These "shoulds" are scientific in the sense that they reflect discovered process-product relationships. The activities have all been found to increase achievement in primary programs.

I think a couple of observations are in order. First, from my perspective, I think we are better off to have this knowledge than not. However, it is very limited and level specific, and there is a great deal of research yet to be done before we will have enough real data to evaluate the scientific basis of teaching. This stance comes awfully close to espousal of the position that "what works is good." There still needs to be moral, aesthetic, and ethical criteria before a relationship becomes a "should." What if, for example, we found that by frightening students, higher achievement would result, *would we still do it*? Finally, there are professional criticisms of the product-process model. Doyle (1978), for example, asserts that "reviewers have concluded with remarkable regularity, that few consistent relationships between teaching variables and effectiveness criteria can be established."(p. 69) I cite this to remind us that research has its honest critics. We are far from unanimous agreement that a scientific basis for teaching is obtainable at present.

The Problems of Value In Evaluation Of Teaching

The crux of the problem in evaluating teaching lies in identifying criteria that define "good teaching." We have just discussed the attempt to determine values empirically, through process-product relationships. Central to this process, however, is the value assumption that "good" teaching results in increased pupil achievement.

But short of this, and we are quite short of it, there are a host of other considerations that come to mind. Achievement, for example, is related to learning attitudes, values, appreciations, moral principles, cognitive perspectives and other outcomes. At present we have only minimal ways of measuring these out-

comes, to say nothing of connecting them empirically to teacher behaviors.

Christopher Jencks (1972) and others have forcefully argued that the school may be best able to develop attitudes and values rather than cognitive achievement. If this is so, then we must pay a great deal of attention to what teachers are doing, i.e., how they are teaching attitudes and values in areas which, for example, lead to such outcomes as: growth in student confidence, student belief that doing school tasks has worth, student persistence in working, and student self-management and direction.

The basic questions are what actions do we value on the part of the teacher, and why do we value them? Do we value formative actions on the part of the teacher in contrast to summative outcomes like achievement of all sorts? Then, for example, are teacher plans, teacher tests, or teacher selection of materials bases for judging good teaching? In Phillip Jackson's (1968) terms do we judge the preactive decisions teachers have made as well as the interactive processes they engage in?

The Complexity Of Evaluating Teaching

The evaluation of teaching makes sense in a rather simple, technical framework if, for example, we assume that pupil learning is the result of teacher behaviors or if we assume that teachers are free to choose any behaviors. Now, we all know these assumptions are not true; teaching is a much more complex situation. To illustrate the complexity of the task, let us look briefly at some of the factors involved in teaching. Lundgren and his Swedish colleagues (1979) recently published a study of "Code, Context, and Curriculum Processes." Essentially, what they demonstrate with a variety of studies are factors that *restrain, govern,* and *regulate* the teaching process. They identify three groups of systems that they found are linked to achievement: the curriculum, the administrative apparatus, and school laws and school legislation. They point out that in the teaching situation these factors cannot be influenced by either the teacher or the students. The only way of influencing them is through political processes. Lundgren says: "The main argument is that, to explain any educational process, we must have a conceptual apparatus that relates the economic and social structure of society to the teaching process...." (Lundgren, p. 33). I know of *no evaluation of teaching material*

which takes into account the social and economic structure of society. In essence, Lundgren is commenting on something we have all experienced in teaching at whatever level. When, for example, the curriculum specifies objectives and materials; when administrative rules and regulations specify the use of time and space (scheduling); and when formal rules or regulations come from the state, there is much less flexibility in teaching than evaluation procedures assume.

Stop and think! When your goals, the length of the periods, the use of places to learn, and kinds of acceptable activities are predetermined, "good" teaching may well be more of a judgment about the *conformity* of the teacher than a demonstration of his/her skill. This should be enough to give pause, but, not so! As I'm sure you are aware, there is almost complete unanimity among thinkers and researchers about two points. Good teaching is teaching that increases student learning. And, no more than 30%, at most, of the differences in achievement can be explained by schooling practices.

Of that 30%, the largest factor in the school seems to be the characteristics of other students attending the school. If, for example, a student is a low achiever with average potential, when he/she goes to school with high achievers in the majority, he/she begins to take on the values and behaviors of the "cognitive majority." Teacher behavior or program alternations seem to have less impact on achievement, although some impact shows up. This seems to say that factors such as socioeconomic background and intelligence are far more important than teaching processes *per se* in influencing the amount of learning that takes place in school.

Gerhard Afrwedson (1979) remarks that in teaching:

> There are no such things as general and optimal "personalities," "roles," or "methods" in the teaching profession, and the reason is very simple: The contextual setting of teaching situations varies in so many different dimensions and directions that what is a "good" personality, role or method in one situation may be almost the "worst" in another situation. (p. 88)

The regularity that exists (and some certainly does) is not due so much to the teacher, but rather to *common contexts*.

There is evidence that there is little hope of changing schools or teaching/learning by changing factors *within* the school, i.e., factors such as methods, equipment, teacher roles or personalities.

Further data raising questions about the power of teaching comes from anthropologist Dorothy Lee (1976), who discusses the culture of the Eastern European Jews. She says that a strong value supporting the pursuit of learning was absolutely necessary if there was to be education at all because, according to our theories, everything was wrong with the educational system, poor timing (in terms of development), poor curriculum, and poor teaching. In most cases, she says, the teachers in the schools are described as ignorant, miserable, unjust, and cruel to children:

> Sholern Aleichern reports that whippings were so usual in his school that there was no feeling of shame attached to them; the resulting sores were painful, but the boys said, 'they'll heal before we are married.' The teacher is generally described as having taken up teaching through default; fallen into his profession because he failed elsewhere. (p. 55)

Yet, out of these miserable classrooms with poor teachers came numerous famous scholars characterized by a burning, self-directed application to study which resulted in magnificent contributions to western civilization.

How did this love of learning, this personal search survive the miserable classroom? In the Jewish community and culture, learning was presented as of infinite value. In addition it earned one true respect and cultural access into adult life and the fellowship of the community. Lee says, "...to my mind the secret (or success of education among Eastern Jews) lies in the supporting values of the home and the community, as well as in the questioning mind...." (p. 60)

Essentially, Lee proposes a similar conclusion as the social research studies of Coleman (1966), Jencks (1972), and others. The conclusions seem to be that the values and attitudes of the family and community are far more important than the behavior of teachers in affecting student achievement in the schools. Thus, the evaluation of teaching is bedeviled on one hand by the "frame" factors that prescribe, govern, and constrain teacher behavior, and, on the other hand, by external factors that influence the achievement of learners. It cannot be assumed that teachers have much control over the achievement outcomes of a student. Thus, how *just* are evaluations that are based on examination of process-product relationships?

A Philosophical Quandary

Fundamentally, there are certain philosophical problems associated with the logic of the evaluation of teaching. If it is assumed that teacher behavior causes student learning (a cause-effect model), then teacher behavior is caused by other factors also. Thus the teacher "can do no other," or has no choice in his/her own behavior. Evaluation of teaching then becomes judgment about the causative factors that condition what a teacher does, not a judgment about the behavior of the teacher.

On the other hand, if we see the teacher as a purposive, choosing, decision-maker, then teachers are responsible for their teaching behavior. From such a vantage, they *could* act differently and are responsible for their teaching, that is, their human actions as teachers. Evaluation of teaching seems to be more just under this set of assumptions. Yet this, too, is deceiving. For if we attribute purposiveness, choice, and decision making to the teacher, we must attribute these same characteristics to students, as human beings. In this case what students learn is certainly strongly influenced by their purposes, personal choices, and their own decision-making process, which in many ways are beyond the influence of the teacher.

We are thus in a quandry. The mechanical model of cause-effect fits the dominant logic of the evaluation of teaching, yet it breaks down because, in terms of mechanistic logic, there is no room for concepts such as responsibility or accountability. Who, for example, is accountable for teacher behavior? How far back in history, or how far out in society, must we go for the "culprits"? Which sources of data do we seek and accept? Do we assume that teachers act the way they do because their behavior is caused; or do we assume that teachers act the way they do for the reasons they give for their actions? Or both? And, how do we account for both in an evaluation of teaching?

The Problem of Teaching

Not only is evaluating teaching effectiveness problematic, but teaching itself may be problematic. The major way that teacher effectiveness can be established is by relating it to student learning or achievement. Thus, more effective teaching is expected to result in greater student achievement. With this in mind, we are able to distinguish some teachers who are more effective than others. So far, so good.

What is truly doubtful is whether teachers *per se* have much of an impact on student achievement. Teachers who do significantly better may very well not contribute much to the larger scheme of things. In other words, teaching as we know it may not be a very effective way of facilitating achievement. Concern for teaching may be much ado about very little!

I am reminded of Illich (1970) and Reimers' (1971) concern for learning networks at this point. Reimer put it this way:

> Learning a skill, learning to practice it with someone else who is learning it, and learning how others have learned it, are three different things... Schools try to find teachers who combine all three kinds of learning, but understandably they often fail. (p. 128)

Reimer proposes that priorities in regard to educational resources be changed. First, attention should be paid to the availability of information in the form of records, the instruments that produce and interpret these records, and other objects in which information is stored. Second, priority must be given to the availability of skill models, i.e. people who can demonstrate the skills. Third, priority must go to the availability of real peers, fellow learners with whom to share learning. And, last priority should go to providing educators, who by virtue of experience, can facilitate the use of learning resources.

Thus, teachers *should not merely give information* for they are probably least effective of all the media available for this purpose. Information lies in resources and in a wide variety of role models as well as in practice with learning peers. Educators can best make "matches." Students will need help in selecting learning resources, choosing skill models, discovering peers, and finding leadership in difficult endeavors. The idea is to get the teacher out of the way of the learner and learning.

I suspect that teachers, as we now know them, will be with us for a long time. However, it should at least be noted that the evaluation of teaching may be premised on an assumption of the significance and importance of "teaching" that is not demonstrable.

The Problem Of Evaluating

The process of evaluating teachers is also problematic (Macdonald, 1974). Evaluating is an act of faith on the part of the evaluator. Evaluators have faith in the value of evaluation. The

evaluation process itself is rarely, if ever, critiqued in terms of criteria such as:

1. *Meaning of Evaluation Behavior in the Context.* Is it neutral or does it serve certain specific interests? In whose interest is evaluation?

2. *Empiricality.* Do we have empirical data showing that teaching is effective?

3. *Validation of Ends.* Do we know that the results of evaluation are what we want? Is there any follow-up to see that the "life influence" decisions (pay, tenure, retention, promotion) are worth it in terms of how little we know about evaluation?

4. *Unintended Consequences.* How do we account for declining teacher morale, changes in situations because they are evaluated, the conformity of curriculum coming from evaluation, or the loss of creative teaching?

Conclusions

In effect the whole business of evaluating teaching is problematic. There are very real philosophical, analytical, empirical, and social problems that suggest going very slowly. Unfortunately, we seem to be buffeted by forces that do not recognize this complexity, and thus we must assure a difficult, if not impossible, responsibility.

References

Arfwedson, Gehard. "Teachers' Work," in Lundgren & Petterson, eds., *Code, Context and Curriculum Processes.* Stockholm: Stockholm Institute of Education, Department of Educational Research, 1979.

Coleman, James, et. al. *Equality of Educational Opportunity,* vol. 2. Washington, D.C.: U.S. Office of Education, 1966.

Doyle, W. "Paradigms for Research on Teacher Effectiveness," in L. S. Schulman, ed., *Review of Research in Education,* vol. 5. Ithaca, Ill.: F. E. Peacock, 1978, pp. 69-74.

Eisner, Elliot. *The Education Imagination.* Berkeley, Calif.: McCutchan, 1979.

Gage, Nathaniel. *The Scientific Basis for the Art of Teaching.* New York: Teachers College Press, 1977.

Illich, Ivan. *Deschooling Society*. New York: Harper & Row, 1971.

Jackson, Phillip. *Life in Classrooms*. New York: Holt, Rinehart & Winston, 1968.

Jencks, Christopher. *Inequality: A Reassessment of the Effect of the Family and Schooling in America*. New York: Basic Books, 1972.

Lee, Dorothy. *Valuing the Self*. Englewood Cliffs, N.J.: Prentice Hall, 1976.

Lundgren, Ulf and Petterson, Sten. *Code, Context and Curriculum Process*. Stockholm: Stockholm Institute of Education, Department of Educational Research, 1979.

Macdonald, James B. "An Evaluation of Evaluation," *Urban Review*, January 1974, pp. 3-15.

Medley, D. M. and Mitzell, H. E. "Some Behavioral Correlates of Teacher Effectiveness," *Journal of Educational Psychology*, vol. 50, 1959, pp. 239-246.

Reimer, Everett. *School is Dead*. New York: Doubleday, 1971.

Rosenhine, Barak. *Teaching Behaviors and Student Achievement*. London: National Foundation for Educational Research in England and Wales, 1971.

NSPER: 79

Milwaukee Presentation

DALE L. BOLTON
University of Washington

Planning for the Evaluation of Teaching: Guidelines and Principles

Dale L. Bolton

Planning for the evaluation of teaching may be conducted in a variety of ways. The following discussion includes general ideas about planning that may apply to a variety of situations. Group and individual activities are suggested that facilitate the planning process, and ideas for conducting conferences between evaluators and evaluatees are considered.

General Ideas Regarding Planning

Planning is basically a process of making decisions prior to taking action in regard to an enterprise or program. The process can be described as a set of activities, for example:

- analysis of background and context of the enterprise
- establishment of objectives (outcomes desired on the basis of the background information and values of the decision makers)
- prescription and sequencing of activities needed to accomplish the objectives
- identification of people to be engaged in the activities
- allocation of time and other resources needed
- determination of means for assessing success (both intermediate and terminal).

In general one of two approaches may be used in the planning process, goal orientation or problem orientation. The goal orientation approach uses a method of "ideal approximation." For example, in planning for the evaluation of teaching, one would conceptualize an ideal teacher evaluation situation. Constraints would then be examined (e.g., resources and resistances). Then, the ideal would be modified as suggested by the constraints identified, but only when absolutely necessary, maintaining as close an approximation to the ideal as possible.

The problem orientation approach uses a method of "problem reduction" which differs from that of "ideal approximation." Here one would examine the discrepancy between what *should* be happening and what *is* happening, determine potential causes for the problem, eliminate causes, and reduce all of the

problems encountered in the teacher evaluation to as near zero as possible. The difficulty with the problem orientation method is that the problems in a *poorly conceived* teacher evaluation plan might be reduced to zero and it could still be a poorly conceived plan—i.e., in terms of benefiting the teachers or the organization.

This does not mean, however, that a problem orientation should never be used in planning. It does mean that one needs to exercise good judgment in determining which method is appropriate. Sometimes a system needs to be examined in terms of whether it is approximating an ideal system; at other times, one will have to solve problems which arise in the implementation and maintenance of a given system. Therefore, a manager needs to be able to use both approaches.

In planning an evaluation system, one is concerned with both conformity to organizational requirements and adjustment to individual needs and interests. Conformity, for example, may require checking roll daily in all classes or treating children kindly. Such requirements may necessitate development of job descriptions or summary report devices for use in an evaluation system at the end of the year. In any case, conformity requirements demand common standards and criteria for a set of people. These standards and criteria, and their measurement, need to be understood by all involved.

Adjustment to individual needs and concerns involves attention to individual goals and objectives. Usually these are established by or with the teacher, although some may be specified by the evaluator, especially if problems have been noted in the past. These individual goals and objectives should be measurable, worthwhile, understood by all concerned, and approved by the evaluator after a discussion between evaluator and evaluatee. This adjustment to individual concerns and needs is usually accomplished by some type of management-by-objectives procedure accompanied by ad hoc data collection devices.

In planning for evaluation, either with groups or individuals, one should communicate consistently by using precise terminology. For example, certain terms are rather general and can mean any number of things. Goals or purposes may mean any of the following five things, yet each is different from the other:

- mission of a school system
- specific objectives of teacher related to the mission
- specific objectives of teacher related to career goals

- purposes of the evaluation system
- evaluator objectives

Not only are the five items different, they are considered at different stages of the planning process, and people have different responsibilities in relation to them.

As teachers and evaluators consider objectives related to the mission of a school system, priorities must be set. The teacher may state an extensive list of general goals and objectives, outcomes desired, and activities needed to produce the outcomes. To these, the evaluator may add other objectives, thus making an almost overwhelming list. There may be no possible way of accomplishing all of those objectives in a given time period. If the unrealistic expectation of doing so persists, a great deal of stress may be experienced. In order to prevent stress symptoms in both the teacher and the evaluator, establishing priorities among the objectives is essential. It may be necessary either to reduce the objectives to a manageable number or to place them in a priority order when it is clear that all of them cannot be accomplished. These priorities may be changed from time to time. The advantage of the cyclical nature of the evaluation process is that it allows changes in priorities as additional information is gathered and analyzed.

Group Activities in Planning for Evaluation

Certain types of group activities may be beneficial in planning for the evaluation of teaching: general faculty meetings, department or divisional meetings, grade (or other level) meetings, and team meetings. The purpose of such meetings is to increase efficiency by dealing with certain topics with the affected groups. To discuss all topics on an individual basis would consume more time than the group meetings.

Two types of activities usually occur in group sessions: 1) clarification processes, and 2) actions related to work production. Clarification generally relates to organizational goals, purposes of evaluation, and context of work (situational analysis), which may include a review of job descriptions or an explanation of summary report evaluation forms, group and individual tasks to be performed, and time lines (explanation of when events will occur). On the other hand, actions related to work can include such things as the writing of work unit goals (e.g., departmental or team goal statements), discussion of the context or conditions of work (including detailed discussion of the

nature of students and resources available), division of labor among members of the work group, and orientation of new people to details of the job.

Responsibilities of the Evaluator

Both the evaluator and the evaluatee should engage in planning activities prior to meeting in a conference. Generally, in planning for the conference, the *evaluator* should have the following three goals for meeting with the teacher:

1. *Clarification.* This includes a review (in order to make sure the teacher understands the topics presented during the group activities) of district and building goals, a review of the purposes of evaluation (both the general purposes of the school district plus those that particularly pertain to the teacher), and a review of the context of the teacher's assignment (including the job description, the summary evaluation form, and any unique aspects of the assignment relating to students, resources, or constraints).
2. *Review of teacher's plan.* This includes a review of the goals and objectives, strategies and procedures, and means of measurement.
3. *Action regarding the teacher plan.* The evaluator has responsibility for approving or disapproving goals and objectives established by the teacher. Other actions include deciding what measurement will be used to determine if objectives have been met, determining the sequence of next steps in the evaluation process, and approving, or disapproving, resources requested by the teacher.

If these goals are to be accomplished, the evaluator must bring pertinent information to the planning conference. First, the evaluator should review and bring to the conference information regarding the teacher, such as: length of teacher experience, previously identified strengths and weaknesses, general level of maturity, and whether probationary status is being considered.

Second, information related to the position or assignment of the teacher should be considered. This may be clarified by means of a position analysis or job description. Various formats can be used for these descriptions, as illustrated by the outlines provided in Outlines 1 and 2. The evaluator may request that each teacher complete such an outline in order to

review the details of the position. In other instances, descriptions may be kept on file and merely reviewed or modified periodically. At any rate, the evaluator may desire to review and bring to the conference information of this kind.

Outline 1
Position Analysis

I. Nature of the position
   A. Broad goals of the system
   B. Organizational structure
   C. General expectations
      1. In the classroom
      2. In relationships with members of the organization
      3. In relationships with external groups or individuals

II. Static and dynamic features of the position
   A. What is likely to be important at the beginning of the year, and later?
   B. What is likely to change?
   C. What is likely to influence this change?
   D. How will this position be affected by other people?
   E. How readily does the community accept change?
   F. What is the nature of the students in the class?

III. Teacher behaviors (processes)
   A. Required/desired
      1. In the classroom
      2. In relationships with members of the organization
      3. In relationships with external groups or individuals

IV. Student outcomes (products)
   A. Required/desired
      1. Achievement and other cognitive skills
      2. Attitudes, social skills, other effective attainments
      3. Psychomotor skills

V. Teacher characteristics
   A. Aptitudes
   B. Skills
   C. Social requirements
   D. Interests
   E. Physical traits

Outline 2

Job Description
(including school district common criteria)

I. Position detail
   A. Grade level
   B. Subject area(s)
   C. Daily assignment
   D. Extra class duties (compensated, noncompensated)
   E. Leadership responsibilities (e.g., department head, team leader)
   F. Anticipated changes

II. Student outcomes desired
   A. Academic development
   B. Social development
   C. Psychomotor development
   D. Attitudes and interest development
   E. Vocational skills development
   F. Remedial development
   G. Other....

III. Instructional processes desired
   A. Individualization
   B. Presentation
   C. Discussion
   D. Demonstration
   E. Experiments
   F. Use of audiovisual equipment
   G. Small groups
   H. Role playing, simulation, games
   I. Other....

IV. Teacher characteristics sought
   A. Flexibility

B. Commitment to accomplishment
C. Enthusiasm
D. Clarity of expression
E. Scholarship and conceptual skills
F. Relating to students
G. Professional orientation
H. Modeling appropriate behavior
I. Other....

Third, the evaluator may want to review various ways of measuring results of teaching and bring certain information-collecting devices to the conference for review by the teacher and the evaluator. The devices should be pertinent to the objectives established by the teacher, which implies that the teacher has submitted written objectives to the evaluator prior to the conference.

In addition to thinking about the types of information needed for a successful conference, the evaluator should plan for the establishment of a climate for the conference. Some thought should be given to the types of communication to be used, types based on the background information available about the teacher and the position. For example, should the conference emphasize telling, asking, exchanging, demanding, or encouraging? The evaluator should consider the point of view to be demonstrated, such as mutual goal setting, facilitating teacher reaction, exchanging views, degree of involvement in decision making, and reaching agreement on methods and resources. Above all, the point of view should clarify the type of leader/subordinate relationship desired. The appropriate setting can also affect the climate of the conference whether it is in the evaluator's office, the classroom (where use of space and materials can be seen), or in a neutral setting such as a teacher workroom. Finally, climate will be affected by the setting aside of an adequate amount of uninterrupted time for the conference.

Who should participate in the planning conference is another consideration for the evaluator. The conference can be simply a two-way interaction between the evaluator and the evaluatee. In certain circumstances, however, a third party, such as a department chairperson or district curriculum specialist, may be included. In other instances, peers or professional organization representatives may be involved.

Finally, the evaluator should give some thought to the way the conference will be terminated. Regardless of the content of

the conference, it seems advisable to include in the closing comments some summarizing statements about whatever conclusions and agreements were reached, as well as next steps to be taken by both the evaluator and the evaluatee. The closing should also include a reminder of the next evaluation meeting between evaluator and evaluatee.

Responsibilities of the Evaluatee

The primary responsibilities of the *evaluatee* in planning for evaluation relate to the position itself and the goals and objectives associated with the position. Regarding the position, the evaluatee should consider first the expectations others have for the assignment. In what way do other people expect the teacher to conform to organizational requirements? Likewise, in what way do others expect the teacher to develop individualized activities and outcomes? Some positions require teachers to follow rather rigid curriculum guides, whereas others are predicated on creativity and unique behavior. The teacher should take these expectations into consideration when planning.

Second, the teacher should analyze personal expectations for performance. This analysis may include such things as the evaluatee's long-range career plans (and how this position relates to the longer-range plans), previously identified strengths and weaknesses, the desire to influence change in the organization, the need to coordinate activities with other individuals, and a desire for support of planned activities.

Third, the teacher should consider personal status in the organization, either a "retain" category or a "consider for release" category. Goals and activities will be focused differently in these two categories, and naturally the teacher's plans will be affected differently. Regardless of status, the teacher's role will include: 1) conceiving objectives and procedures relevant to the objectives and, 2) identifying resources desired and required to carry out the procedures.

In preparation for writing goals and objectives, the teacher should think about mission in relation to the total organization, cost to accomplish the objectives, and likely results. Otherwise, one is likely to end up with a list of routine activities. Although all three of these items should be considered, one should focus on results first. Specification of what is desired in terms of final outcomes should precede identification of activities so that the

activities designed will be appropriate to the outcome desired.

In writing objectives the level of generality is of some concern. These statements should be helpful to both the evaluator and the evaluatee. Objectives should be specific enough to give direction and to facilitate judgments regarding progress and excellence. However, specification should not be so detailed that it causes cumbersome paperwork, results in a lack of attention to important tasks, tends to stifle creative and assertive people, and/or attracts only the passive and conforming people to the job.

Two general criteria should be used in examining objectives. First, they should be worth doing! This means they should be examined to determine if they are realistic and likely to be attainable, relevant and important to the position, challenging beyond the routine tasks expected of all teachers, and consistent with the organization's general thrust. Second, objectives should be well written. Objectives should specify the outcome or result expected, the action to be taken, who takes the action, when the action should be taken, the cost, and the conditions needed for the objectives to be accomplished. In addition, the objectives should clearly state the way that the results are to be verified.

The Planning Conference

Once evaluator and evaluatee have done their individual planning, the two (with others, if desired) should complete the plans together. The planning conference should begin with a review of the agenda, including specification of topics to be considered, sequence of the topics, and decisions to be made. Although the evaluator is responsible for developing an initial agenda, there should be opportunity for the evaluatee to add items to be discussed.

Both should concentrate on clarity of communication throughout the conference by periodically thinking about the following questions:

1. Are the purposes of the evaluation clear?
2. Does the other person "understand" the job?
3. Are goals and objectives worth doing? Are they written clearly?
4. Have resources been discussed thoroughly and agreed upon?
5. Do oral agreements need to be converted to written form?

6. Is the responsibility of each person for the next step clearly understood?

One of the major tasks during the planning conference is to make decisions about measurement. Certain questions need to be answered. First, how will information be collected? Basically, there are three ways that information can be collected in the behavioral sciences: by observation, by asking questions, and by examining written records. Most of the information collected for evaluating teaching is done via observation (either of students, teacher, or environment for learning) or by asking questions (via tests or by self-report devices completed by students). Agreement should be reached regarding what information will be measured by each of these methods.

The second question is what is to be measured? One may be concerned with processes, products, or characteristics. Processes may include such things as presenting ideas, organizing groups, grading papers, or organizing space and time. Products may include changes in students, or student and parent satisfaction. Characteristics may include such factors as reliability, appearance, or sociability. No case is made here for any of these variables. But, an appropriate procedure would determine the variables to be used locally so that both the evaluator and the evaluatee understand what is to be measured.

Third, there is the question of who will do the measuring. Some measurement is likely to be done by the evaluatee in a self-evaluation process, while other measurement will be done by the evaluatee cooperatively with the evaluator. Still other measurement will be done by someone other than the evaluatee. During the planning conference, evaluator and evaluatee should agree on who should be collecting what information.

Fourth, when and where will the measurement occur: during or after instruction, by the teacher or by others? Generally, measurement occurs in relation to an event. For example, in a classroom discussion observed by an outsider, measurement of the verbal interaction occurs during the event. In another case a teacher might use a tape recorder during the event and then measure certain aspects of the interaction following the discussion. Likewise, the teacher may choose to measure the results of a discussion by giving the students a questionnaire to complete after the event. Events, behaviors, or results of behavior may be measured in the classroom or outside the classroom. Where objectives concern the process of teaching, measurement is likely to occur in the classroom. Where the objectives concern the performance of the teacher in nonclassroom activities

(i.e., functioning as a member of a curriculum-development committee, contributing to departmental planning, etc.), the collection of data is likely to occur outside the classroom.

Finally, what types of forms will be used? Evaluator and evaluatee should differentiate between process data collection forms and summary report forms. Quite often school systems develop a single summary report form for use in the district and then attempt to use this summary form for collecting process information. This is done in spite of the fact that such forms have not been designed for the specific measurement purposes. Data collection forms may be developed by people other than the evaluator or evaluatee, but care should be taken that they are appropriate for the objectives agreed upon. Where no form is available, both the evaluatee and the evaluator should be creative in designing forms that facilitate accurate measurement of the factors selected.

During the planning conference, the teacher and evaluator should discuss division of responsibility between them for parts of the evaluation process. Each teacher has a responsibility for establishing objectives that are compatible with and contribute to the overall mission and goals of the school district. This is not the responsibility of the evaluator, although the evaluator may make some suggestions. Likewise, the district may choose to establish common objectives for a group of teachers. The objectives of the teacher are not synonymous with the goals of the district or building. The processes and products of the teacher are unique to that teacher's assigned responsibilities.

On the other hand, the evaluator has the responsibility for approval of the teacher's objectives. When approved, it is assumed that the objectives are compatible with building and district goals. Both the evaluator and the teacher should understand that establishing goals and objectives means commitment of time and effort; approval and acceptance means support via school district resources.

Finally, a warning is appropriate regarding planning for the collection of information. Teachers and evaluators should be reminded that the definition of evaluation does not include a judgment concerning the personal worth of an individual; rather, the concern is with performance and the accomplishment of objectives. This focus is essential to good planning for the evaluation of teaching.

The Bases For Evaluating Teaching: Philosophy, Context, and Purpose

Dale L. Bolton

Three significant philosophical aspects of evaluating teaching need to be emphasized. They are: 1) definitions of philosophy and evaluation, 2) some assumptions relating to evaluation, and 3) elements needed for a system of evaluation based on the assumptions stated. The context of evaluation, i.e., the nature of the environment within which evaluation occurs, may be described as having some, or all, of the following characteristics: uncertainty, complexity, professional/union tension, accountability, and a perceived scarcity of time. The section on purposes of evaluation will include a discussion of some issues as well as an identification of some specific purposes.

Philosophy

Philosophy can be defined in at least three ways:

1. A science which investigates the facts and principles of reality, human nature, and conduct
2. A body of principles underlying a given branch of learning or major discipline
3. An activity that raises prior questions regarding knowledge, morality, aesthetics, and social issues.

This discussion will neither deal exhaustively with the science that investigates facts and principles of reality, nor will it set forth a complete body of principles. However, definitions will be proposed and assumptions stated which, by the way they are stated, will raise questions about our knowledge bases and the social issues with which we are involved.

Evaluation involves more than measurement. Measurement is merely the quantification or quasi-quantification of events, behavior, or results of behavior. It is sometimes called assessment. Evaluation is also the making of judgments regarding events, behavior, or results of behavior in light of predetermined and well-understood objectives. Underlying these definitions are two presuppositions. One is a careful differentiation

between measurement and evaluation. Many people have considerable difficulty measuring what they desire to measure, but it would be a mistake to say that they are having difficulty in evaluation. They are having difficulty with a *part* of evaluation, but not with its entirety. Measurement and evaluation should not be used as synonyms.

Another presupposition is that evaluation should involve predetermined objectives. Unless predetermined objectives are involved, it becomes a matter of shooting at a target and then drawing a circle and saying, "Yes, I hit it." But such a process also allows someone else (an evaluator) to draw a circle in another location and say, "See, you didn't hit it." Not only can the target be in a different location from the spot being fired upon, but it can be completely out of sight. The reasons for predetermined objectives are that they provide focus for teaching activities, and they prevent capriciousness on the part of the evaluator and the evaluatee.

There are several basic assumptions underlying the philosophy of evaluation of teaching. First, it is assumed that evaluation of teaching does occur. All teachers and their teaching are evaluated; they are evaluated now, they have been evaluated in the past, and they will be evaluated in the future. That is not the issue. The real issues revolve around determination of the criteria to be used and the soundness of data to be collected so that a systematic evaluation process can be established.

A second assumption is related to the conditions under which evaluation should occur. For one thing, good evaluation is facilitated when those involved in the process exhibit respect for the worth and dignity of individuals. This basic position of mutual respect is needed in order to deal with the reality of the situation; otherwise, either party may engage in gamesmanship. Second, there must be a sensitivity to the fact that the person being evaluated feels a kind of inner tension between wanting to improve and not wanting to change behavior. This inner tension sometimes exhibits itself in inconsistent behavior; initial eagerness may turn to hesitancy or resistance, and agreement may change to questioning. Third, there should be an awareness of differences in expectations about the meaning of the evaluation process. Some evaluators expect high productivity and evaluate in terms of these expectations. Some evaluatees expect evaluation to be a faultfinding process, or completion of a required form, or an emphasis on comparisons with others. Fourth, although the individual being evaluated is

a member of a profession of peers, teaching occurs within an organizational heirarchy. Until recently, teachers and administrators belonged to a unified professional organization. However, with the advent of negotiations, this is no longer the case in most school systems. In addition, it is a fact that the responsibilities of positional leaders in any organization differ from the responsibilities of nonpositional leaders. Finally, it is assumed that good evaluation presupposes a clear recognition of the constraints placed on people and upon the system of evaluation. There are constraints of resources, time, resistance by individuals, and legal requirements in most school systems. All of these constraints have an impact on the process of evaluation itself.

A third type of general assumption involves the concept of commitment. For an evaluation system to function optimally, those who are involved must be committed to two things: 1) excellent results, and 2) satisfaction with the work being done. If a teacher is not committed to these two things, it is very difficult for the evaluator to be highly motivated to work with the teacher. If the evaluator is committed only to mediocrity—and perhaps to satisfaction with the life *off* the job—it is difficult for the teacher to become very enthusiastic about evaluation. Some administrators are keenly aware of the fact that evaluation allows them to interact with teachers about things that are really important to both themselves and teachers. Therefore, they find evaluation not only personally satisfying, but also satisfying to the teachers. Yet this situation is dependent upon the two parties' commitment to excellent results and satisfaction with their work.

Who is involved in the process of evaluation is a fourth major assumption. It is assumed that a good system of evaluation includes not only self-evaluation but also evaluation by another person. Further, it is desirable that the external evaluator be responsible to the organization, since the evaluatee has the option of ignoring data and advice provided by peers. Peer evaluation may be beneficial for certain purposes, but it is essential to have at least one external evaluator to prevent teachers from ignoring feedback and criteria for evaluation established by the organization. In addition, peer evaluation cannot perform the function of approving the goals and objectives of the evaluatee, since to do so would negate the definition of a peer.

Finally, it is assumed that administrators are involved in the evaluation of teaching. They have responsibility for examining strategies for increasing and sustaining the productivity of the

organization and increasing and sustaining the satisfaction of teachers and staff. The general strategy for increasing the productivity of individuals has to do with setting goals at a higher level than current productivity, yet within reason. If individuals do not set and strive for higher goals, increased productivity will occur only by chance. The major strategy for sustaining that productivity is to provide feedback as to whether or not performance is at the desired level.[1] The strategy for increasing morale and satisfaction with the job has to do with the way communication occurs during the goal-setting and feedback processes. Since goal setting and feedback are integral parts of evaluation, it is reasonable to conclude that an administrator who is concerned with both production and satisfaction should become concerned with designing a functional evaluation system.

Thus far, discussion of the philosophy of evaluation of teaching has dealt with certain definitions and assumptions, each of which raise prior questions. These definitions and assumptions form the bases for identification of elements needed in an evaluation system. To determine whether these elements[2] are present in an existing evaluation system, an administrator might ask the following questions:

1. *Does the district use a process that is cyclical and/or continuous?* The model in Figure 1 illustrates three basic phases: planning for evaluation, collecting information, and using information. Each phase has certain things that must be done, but the important idea here is that once information is collected, it is used to make decisions. Once those decisions are made, the planning is done anew for the next cycle. The model assumes that decisions are to be made on the basis of systematically collected data. Further, the decisions that are made will assist in making plans for the collection and analysis of additional information.

2. *Is a systems analysis perspective used, in that there is an emphasis on input, process, and output (product)?* Does evaluation take into consideration the abilities and motivation of

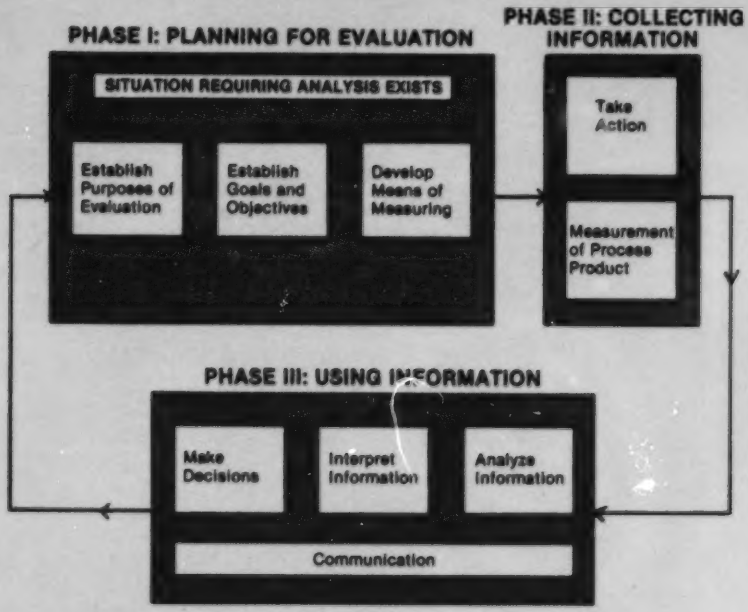
---

[1]The function of feedback, positive or otherwise, is not to encourage people to perform better but to allow them to know whether they are performing in a manner satisfactory to themselves and the organization. Higher aspiration, not positive feedback, is the motivating factor for increased effort and productivity.

[2]Since these elements are discussed in some detail elsewhere, only a brief explanation is provided here. See Dale L. Bolton, *The Evaluation of Administrators in School Systems* (New York: Teachers College Press, 1980).

Figure 1

STAGES IN THE EVALUATION OF SCHOOL PERSONNEL



children being taught; are teaching procedures being examined critically; and, are the results measured carefully? The model in Figure 1 is one type of systems analysis model for evaluation.

3. *Is there a conscious attempt to interrelate the processes and products of various levels of the organization?* Figure 2 illustrates this idea, in that some of the processes and products of teachers, principals, and superintendents are identified. The model indicates that the processes used by the teacher are designed to obtain certain student results. However, the processes used by the principal are *not* designed to obtain student results directly but to bring about a change in teacher behavior.

Figure 2

PROCESS — PRODUCT RELATIONSHIPS

	PROCESS	PRODUCT
TEACHER	Classroom interaction Management of: • Materials, Space • Equipment. . . .	Student Learning: • Cognitive • Affective • Psychomotor
PRINCIPAL *(The Office)*	Teacher — Principal Interaction • Supervision • Inservice Training	Teacher Behavior
SUPER-INTENDENT *(The Office)*	Superintendent — Principal Interaction • Management • Development Program	Principal Behavior

This means that the principal's product is the teacher's process. Therefore, any discussion of process/product relationship must specify which level of the organization is being discussed or confusion will result. Two things appear to be important here.

First, when a person changes positions (e.g., from a teacher to a principal, or from a principal to a superintendent), it is extremely important to make a change in actual role, i.e., in the emphasis on the processes used and products expected. If the superintendent continues to work directly with teachers (because of having learned how to do this as a principal), or if the principal continues to spend most of the time with students (because of having learned how to do this as a teacher), then the wrong role is being performed. Second, unless evaluation permeates the managerial style of the total school system, i.e., unless the process/product relationship is evaluated at all levels of the organization, there is a tendency for people to perform functions they desire to perform rather than those needed for the organization to function properly.

4. *Is the evaluation system designed so that it is related to other subsystems within the district?* Figure 3 illustrates how information flows among the main elements of the personnel system.[3] The model is based on two assumptions. First, the supervision and inservice training program ought to be individualized, based primarily on information from the selection process and the evaluation process. Second, the selection, supervision, inservice training, and evaluation processes have different functions—namely prediction, treatment, and feedback respectively. Therefore, the information from the evaluation process is useful in validating the selection process, in designing an individualized supervision and inservice training program, and in making decisions regarding changes of assignment.

5. *Does the evaluation system provide direction to the individual and the organization?* During the planning phase depicted in Figure 1, direction is provided by clarifying the purposes of the total evaluation system and by setting goals and objectives.

6. *Does the evaluation system provide a planning phase that includes the acquisition of the resources needed to accomplish objectives?* The time to consider resources needed is during a discussion of goals and objectives. Otherwise, unreasonable goals may be set or evaluators may mistakenly assume that available resources are sufficient for accomplishing what is

---

[3]For additional information, see Dale L. Bolton, *Selection and Evaluation of Teachers* (Berkeley, Calif.: McCutchan Publishing Company, 1973), pp. 133-134.

planned. Realism regarding circumstances needed for accomplishment of objectives is as necessary as honesty in communication.

7. *Does the system of evaluation provide for making intermediate decisions?* If the process is cyclical and if the cycle is repeated often, the major decisions being made are frequently intermediate rather than terminal. The intermediate decisions are primarily concerned with whether the goals and objectives were reasonable, whether the processes were fully implemented, and whether the procedures need to be changed.

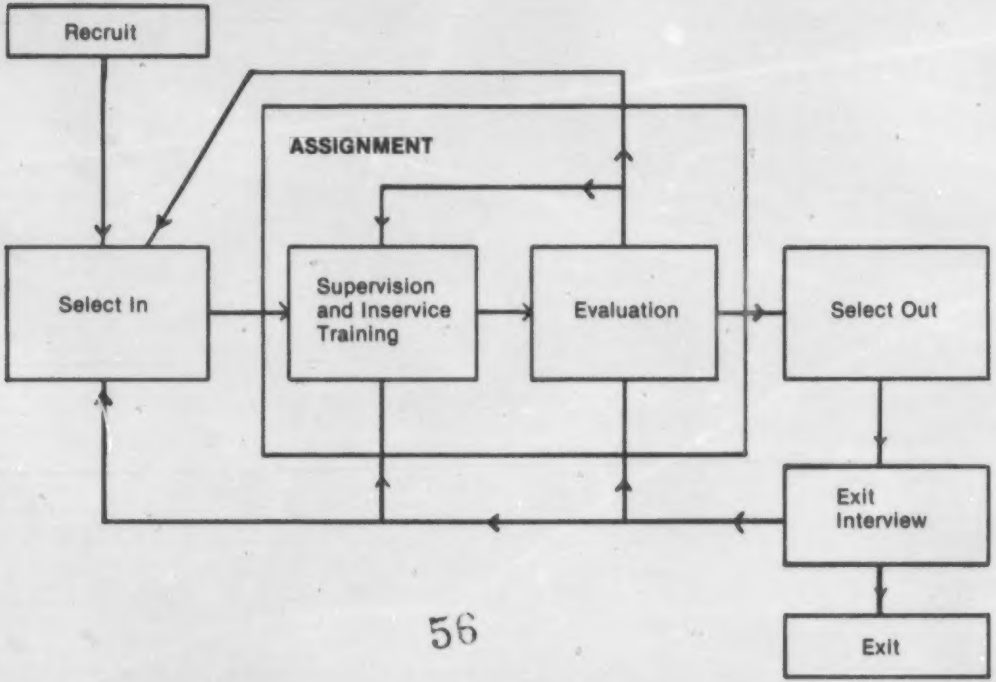
8. *Is the evaluation system designed to facilitate making terminal decisions?* Figure 3 shows information flowing from evaluation to the "select out" phase, illustrating that evaluation can be used for making decisions regarding changes in assignment. Changes in assignment can be made within the district, or individuals can be selected out of the school system. Where the individual initiates this action, a request for transfer or resignation occurs; where the district initiates the action, the individual is transferred or released. In either case, the evaluation system should be designed to provide information to the individual and/or the organization that will facilitate the decision.

9. *Is there evidence that the organization provides support for the process of evaluation?* For evaluation to be successful, it cannot be imposed from the top-level management without full support. This includes top-level reinforcement, integration of evaluation into the total managerial style of the organization, material resources to do the required tasks, encouragement and psychological support from the board of education and the superintendent, training of all involved, and assessment of the evaluation system to allow modification where it is not working well.

10. *Does the evaluation system include a self-evaluation component?* The evaluatee should be involved in analyzing the teaching situation, in writing goals and objectives, in deciding on types of information that will be helpful in making decisions, in analysis and interpretation of data, and in making decisions regarding changes of goals or procedures. Since evaluatees are involved in situations continuously, it is helpful if they learn to do these tasks independently of others.

11. *Does the evaluation system include external evaluators?* Although the external evaluator gets involved in many of the same activities as the evaluatee, a unique function of the external evaluator is to monitor and control the total process. This is exhibited primarily in approving goals and objectives, writing summary reports, and making administrative decisions.

Figure 3 **INFORMATION FLOW IN THE PERSONNEL SYSTEM**



12. *Does the evaluation system incorporate a set of common criteria for those being evaluated?* Common criteria for all teachers in a school district may include processes (e.g., structuring comments such as advance organizers or summarizing statements), characteristics (e.g., clarity of expression), or outcomes (e.g., student enjoyment of school). Those common criteria represent what the local district considers to be important and may be used on summary report forms.

13. *Does the evaluation system include the opportunity for individualized goal setting and individualized criteria?* Generally, teacher-specific goals, processes, characteristics, and outcomes are determined via some type of management by objectives process where teacher and evaluator agree on what makes sense in terms of the particular teaching situation. This individualization is intended to complement the common criteria rather than replace them.

14. *Does the evaluation occur often enough to accomplish the designated purposes?* If evaluation is considered to be a feedback mechanism to help determine whether goals and objectives are being accomplished, it is doubtful if once-per-year feedback (or less) can be very helpful in making decisions regarding goals, implementation of procedures, or effectiveness of procedures. Therefore, it is essential to examine the frequency of evaluation in relation to the purposes for the entire process. The frequency should be sufficient to accomplish these purposes.

15. *Are the purposes for evaluation clearly specified in the written documents regarding evaluation?* This question implies that written documents describing the background, procedures, and forms to be used in the evaluation process are available. In these open documents all purposes should be stated clearly so that no one need wonder why the evaluation system was designed or what its intent is.

Context of Evaluation

There are undoubtedly many ways to describe the nature of the environment within which evaluation occurs. The discussion[4] that follows provides a brief description of five factors that are particularly pertinent to the concerns of evaluators.

---

[4]A more in-depth discussion can be found in Dale L. Bolton, *The Evaluation of Administrators in School Systems* (New York: Teachers College Press, 1980).

Uncertainty

The environment of school districts is uncertain in several ways and for various reasons. For one thing, school districts have become much more politically vulnerable because people have found that they can pressure organizations into doing certain things. Under such circumstances, some political activity has become manipulative. Not only have goals changed because of such activity, procedures for meeting these goals have also changed. The pressures being exerted have come from clients, parents, students, and teachers. The pressures and vulnerability have caused a degree of uncertainty regarding what should be happening in classrooms, and this uncertainty is widespread. However, the amount of uncertainty varies among school districts. Some communities are much more stable and unified in their expectations than others. But where the uncertainty occurs, it is often accompanied by low trust level among teachers, the public, and administrators.

Perhaps the most important element contributing to uncertainty is the fact that we have such diverse populations within the United States and they are expressing different views and value systems. When there is diversity, it is difficult to get agreement about goals and procedures; therefore, there is an unwillingness to work on goals established by others. Combine this with diversity of teacher style, and lack of an accepted set of practices, and the uncertainty is intensified. When vague and diverse goals are being expressed, it is difficult to get people to commit energy to tasks. A person who has one or two specific objectives and is committed to these objectives will exert energy toward accomplishing them. People do not commit energy to generalities.

Complexity

The environment of evaluation is more complex because of changes of expectations of teachers (e.g., expectations regarding the kind of assistance they should have, the kind of role they ought to play, the type of tasks they should perform). In addition, unexpected events in the environment add to its complexity. Principals may develop extensive plans for evaluation in August, yet these plans may be aborted because of unexpected events that interfere. Still another complexity of the environment is related to a changing climate in the area of evaluation

and schools in general. In the past, school environments provided many opportunities for close associations among teachers; now in some schools the separation of teachers approaches a condition of anonymity.

Professionalism/Unionism Tension

The tension evidenced between professionalism and unionism exhibits itself at times as a resistance to leadership—both as resistance to being supervised, and as an unwillingness to serve others as a regular part of one's job. The work ethic is changing among teachers; many want to work a very limited amount of time in order to enjoy activities outside of work. When this attitude prevails, there is little interest in serving others or being supervised in relation to serving others.

One characteristic of a profession is the leeway to define one's role. However, teachers are avoiding the responsibility of individually defining roles; rather, they are using union tactics to negotiate contracts that define collective goals. This may be done in order to prevent the school district from imposing a particular role on the individual members. Regardless of the reason, it causes tensions in the area of evaluation. During the past several years, there has been considerable growth in the power of professional organizations, and along with this increase in power has come an increase in the number of strikes. In general, these strikes have increased tensions between teachers and evaluators.

Accountability

We are living in an accountability era. The management-by-objectives movement that began in industry ushered in a concern for productivity. This same demand for productivity is being exhibited in education. School patrons are expressing an interest in cost/quality relationships. For a long time, taxpayers have expressed concern about absolute costs; more recently, they have begun to examine cost in relation to quality, a more sophisticated concern. Many are aware that increasing costs without increasing productivity contributes to inflation, and this is a major problem in the current economic scene.

Others are bothered by the few teachers who are incompetent or ineffective in the classroom and want the organization to be accountable for doing something about them. In addition, an

increasing desire for accountability has accompanied the recent declining enrollments in some districts with the result that some have indicated a desire to use evaluation rather than seniority on the job as a means of reduction in force.

Finally, some states have laws relating evaluation to accountability in one of two ways. Either they express a concern for effectiveness (another way of stating concern for protecting children from incompetency), or they attempt to protect teachers from capricious administrators who would disallow due process to teachers.

Perceived Scarcity of Time

The perceived scarcity of time is a tremendous problem in administrative circles. Administrators tend to spend their time on what is enjoyable, on what is immediate, on what is causing them tension, on what is required, or on what is important (or perceived as important). An examination of these five things readily reveals that they often conflict with each other. A principal may enjoy (and consider important) visiting classrooms and conducting conferences with teachers, but may be required to do other things in order to meet deadlines imposed by the school district organizations.

In their evaluator role, administrators have a need to manage their time wisely. They must remind themselves that concentration is the price they pay for success. They must concentrate on the results they desire (the kind of productivity they want and the satisfaction they desire to engender in those they evaluate) and the methods they use to accomplish these results.

As administrator-evaluators consider how well their teachers employ time, and how well they accomplish important objectives, certain questions arise: 1) Is the allocation of time sufficient for the task to be accomplished? 2) Is the person using time well? 3) Are unusual and negative circumstances impinging on the situation? 4) Is the person working overtime? If one can answer each of these questions by either a "yes" or a "no" response, an interesting analysis can result.

Table 1 displays the 16 cases that are generated by answering each question either yes or no. An examination of case 1 indicates that if the allocation of time is sufficient, if the person is using time well, if there is an unusual and negative circumstance impinging on the situation, and if the person is working overtime in order to take care of that negative and unusual circumstance, then there is a high probability that the person is

Table 1

Major Considerations Regarding Time and Accomplishment of Important Objectives

A. Is allocation of time sufficient for task?

B. Is person using time well?

C. Are unusual and negative circumstances impinging on the situation?

D. Is the person working overtime?

	Question Responses								Probability of Task Accomplishment				
	A		B		C		D		Definite	Definite	High	Low	Incon-
Case	Yes	No	Yes	No	Yes	No	Yes	No	(+)	(−)			sistent
1	X		X		X		X				X		
2	X		X		X			X				X	
3	X		X			X	X		X	----------	----------	----------	X
4	X		X			X		X	X				
5	X			X	X		X					X	
6	X			X	X			X		X			
7	X			X		X	X				X		
8	X			X		X		X		X			
9		X	X		X		X					X	
10		X	X		X			X		X			
11		X	X			X	X				X		
12		X	X			X		X		X			
13		X		X	X		X					X	
14		X		X	X			X		X			
15		X		X		X	X					X	
16		X		X		X		X		X			

going to get the task done. Case 2 illustrates that if the allocation is sufficient for the task, the person is using time well, negative circumstances come up *but* the person does not work overtime, the probability is low that the task will be accomplished.

An overall examination of the table indicates that there are only two cases where the task is definitely accomplished—cases 3 and 4. In both cases the allocation of time is sufficient for the task, the person uses time well, and there is no unusual and negative circumstance. In case 3 the person works overtime and in case 4 the person does not. In both cases the task will be accomplished, but case 3 presents an interesting inconsistency of behavior. Since the overtime is not needed to accomplish the task, the person is exhibiting workaholic symptoms.

Consider the other two cases where there is a high probability that the task will be accomplished. Case 7 is a situation where the individual is willing to work overtime in order to compensate for using time poorly during regular working hours. Case 11 is one where the individual is willing to work overtime to compensate for a work overload.

In all other cases, work either is not accomplished or has a low probability of being accomplished. What is the point of the analysis? When tasks are not being done, or when consistent overtime is needed to get them done, an individual should examine the situation to see whether the cause is an inadequate allocation of time, a misuse of time, or unusual and negative circumstances. If the negative circumstances are being consistently brought about by the same sources, actions should be taken to prevent this from happening. If the allocation of time is sufficient, negotiations should occur between the evaluator and the district office. If time is being misused, control mechanisms should be initiated to prevent the misuse of time.

Purposes of Evaluation

Before identifying some specific purposes, it may be well to discuss some issues related to purposes.

Issues Related to Purposes

In a district handbook that describes the evaluation process it may be desirable to discuss other issues related to purposes. For example, the purposes of evaluation may be based on beliefs

accepted by people who design the evaluation system, e.g., beliefs about behavior, about the way adults learn, about motivational factors for adult behavior. It may be that certain implied assumptions need to be stated. For example, assumptions are possibly being made about the desire of people to change and improve, about people's commitment to excellence, about responsibility for productivity.

In addition, certain problems may need to be considered. For example, some people think there is only one purpose for evaluation; they express it as "the improvement of instruction." But when the question is raised regarding the meaning of improvement of instruction, satisfactory definition is difficult. So, in addition to facing the question of whether the purpose is single or multiple, there is the problem of how to state purposes so they are acceptable to many people.

A related problem is determining what brings about commitment to the evaluation process. Will teachers be committed to a particular process of evaluation if they are not involved in formulating the process? If not, what is to be done about new teachers in a system? In what way should new teachers be inducted and oriented so they will be as committed to an evaluation process as the people who devised it? These issues and problems are related to specific purposes and may cause individual districts to formulate different evaluation purposes. Although the purposes identified below are not exhaustive, they are worth any district's consideration.

Specific Purposes To Be Considered

Nine purposes are identified which have been discussed in the literature or found in school district manuals of procedures.

1. *Improvement of teaching.* Improvement of teaching may be brought about by changing the performance of the individual, by facilitating the development and growth of the individual, by stimulating self-evaluation, or by changing the conditions of work. It may be easier to change the conditions of work (e.g., by removing certain students from a given room, by grouping teachers, or by providing flexible furniture) than to change the performance of an individual, and the results may be comparable. The point is that one should be aware of various ways of improving teaching as well as changing the behavior of the teacher.

2. *Rewarding superior performance.* Paying people differently is done either to keep people from being ineffective or to

motivate them to be superior performers. The literature is mixed regarding the effectiveness of such practice, and many school districts work hard to reward superior performance by means other than differential pay.

3. *Modifying assignments.* Modification of assignments can be either voluntary or involuntary, within the system or outside the system. Thus there are four distinct types of assignment change: a) a voluntary change within the system would be considered a *transfer*; b) an involuntary change within the system is sometimes called an *honorable transfer*; c) a voluntary change to a position outside the system would be *resignation*; and d) an involuntary change to a position outside the system is a *release*. All of these entail administrative decisions, and many districts require that an evaluation process contribute to the decisions.

4. *Protecting individuals and the organization.* Basically this purpose is designed to protect against unfair criticism, with the assumption that evaluative information can provide a basis for that protection.

5. *Validating the selection process.* If used for this purpose, evaluation provides feedback regarding whether teacher performance matches predictions about performance made at the time of selection.

6. *Satisfying policy, law.* Where school board policy or the state law requires teacher evaluation, this purpose is stated or implied.

7. *Improving decisions.* Several types of decisions can be improved by evaluative information, e.g., broad public decisions such as bond issues and changes in basic curriculum; policy decisions made by board of education such as the emphasis to be placed on various activities or resources; and operational decisions by administrators having to do with such things as modification of objectives or procedures or determining new ways of implementing procedures.

8. *Providing a basis for career planning.* An evaluator may work with an individual so that strengths and weaknesses are understood in relation to opportunities for the future. Some individuals need to be encouraged because of their leadership potential, while others may need to face their weaknesses realistically and plan accordingly.

9. *Contributing to morale and cooperation.* When this purpose is being accomplished, communication will be such that a high level of trust is developed between evaluator and evaluatee.

Concluding Comments

One might reasonably ask whether a given evaluation system is a good one or not. In answering that question, one might ask the following specific questions. Are the evaluative purposes stated clearly and precisely? Is there evidence that the evaluation system is accomplishing its purposes? Is it improving the performance of individuals? Is the system eliminating incompetence by terminating incompetent individuals? Is the system of evaluation cost-effective? Is the system worth the time being invested? Is productivity increasing and satisfaction high?

The background ideas discussed should provide a basis for the planning process for evaluation. There has been an attempt to be realistic in terms of assumptions necessary for describing elements of an evaluation system. The discussion has emphasized the idea that evaluation is a process of making judgments regarding behavior, events, or results of those behaviors in light of predetermined objectives. This means that evaluation is more than a measurement or a rating process, although either or both may be involved. Because of certain aspects of the context of evaluation, real problems of communication and trust relationships evolve. Therefore, careful planning that takes these contexts into consideration is essential.

Finally, it is necessary to approach the evaluation system in the same way one approaches evaluating teaching. One must set goals, collect information, and determine whether the evaluation system's goals have been reached. Only then can reasonable decisions be made concerning the system itself.

NSPER: 79

Albuquerque Presentations

MICHAEL SCRIVEN
University of San Francisco

CAROLYN J. WOOD and PAUL A. POHLAND
University of New Mexico

A Different Approach to Teacher Evaluation

Contributions of Michael Scriven*
at Albuquerque, New Mexico
NSPER: 79

Evaluating staff performance varies in an interesting way between different educational levels. The attempt to use student ratings in staff performance evaluation in elementary and secondary schools will cause unions to scream bloody murder. At the college level they will often scream bloody murder if you *don't* use student ratings. (Sometimes they will scream if you use them, too.) Traditionally, evaluation of faculty performance at the college level has been focused on the faculty member's subject-matter preparation and current expertise, whereas the K-12 level evaluation has been focused on the teacher's achievements in methods courses and his or her presentational style.

Conceptual Issues

A central factor in planning for the evaluation of teaching is a consensus about what is expected from "professional teaching." Such consensus must make a clear differentiation between parental, media, and professional teaching, for example. The essence of professional teaching can be expressed perhaps in the following propositions.

1. Professional teaching must contribute to substantial positive gains in learning, in some or all of content, skills, and attitude.
2. The learning gains from professional teaching should exceed those achieved by reading a book or listening to an audiotape.

---

*_Editor's Note:_ The statement which follows is a summary of the ideas presented by Dr. Michael Scriven to the participants of the Albuquerque session of NSPER: 79. The summary was read and approved by Dr. Scriven. Quotations from this summary to be attributed to Dr. Scriven should be approved by him before publication.

3. And, professional teaching should produce gains that are greater than those produced by a non-professional in the same setting.

In short, if we professional teachers cannot do better than a book, a paraprofessional, or a baby-sitter we are not earning our pay. The gain must be beneficial; something real rather than jargon. Earning 15 unspecified credits in Home and Family Living is a jargon gain. The focal point of professional teaching, and of evaluation in this case, should be on increasing the number of things a student learns to do as a productive member of the home. For example, the individual could have developed the knowledge and skills necessary to prepare meals for the other members of the household, a beneficial and real learning gain.

If the points above are reasonable, it follows that evaluation of teaching, at least in part, involves *comparisons* with the learning that results from contact with a book or baby-sitter. Given a book, and no other assistance, it is logical to assume that on the average the learner will have more knowledge after reading it than he or she had before reading. To give the evaluation of teaching a rational base, that amount of "from-the-book knowledge" needs to be determined. To fail to do so is a failure to estimate the zero point on the scale for measurement of that learning gain for which professional teachers can be credited. The same reasoning holds for the growth that can be attributed to continued interactions with a nonprofessional teacher.

The term "estimate" is used here to recognize variability in the system and that we are working in an actuarial sense. It is not rational to expect all learners to make the same gains from a given book or from contact with a given professional teacher.

A basic question: What is there about teaching professionals that gives them any hope of beating books and baby-sitters as stimulators of learning? A typical "input" answer from workers at the elementary and secondary school levels would focus on preservice preparation and apprenticeship experiences. At the college level the "input" answer would center on the extent of the teacher's knowledge of the subject to be taught. There are two process factors that have to be taken seriously by any teaching professional.

1. *The time factor.* There appears to be a direct relationship between time-on-task and learning that the professional teacher should know about and use in structuring the learners' milieu. If the teacher can effect a conversion of time into learning, the beneficial knowledge gain will be enhanced. This does not mean simply piling on homework *assignments*. Rather, it focuses on

the amount of time the student actually spends as direct learning activity—i.e., as a result of the teacher's efforts.

2. *The respect factor.* Professional teachers recognize that many learners do, or can be brought to, respect teachers and look to them as role models. Professional teachers should use these attributions to stimulate knowledge gains on the part of students.

The studies of *style* approaches to teaching effectiveness—the usual "applied science" perception of teaching—have persistently failed. Despite extensive and varied effort, they have not generated data that would confirm a single teaching style as the best approach—e.g., the zest, sensitivity, and commitment style; the use-the-student's-input style; or the organize-the-materials-for-learning style. Neither has research generated the data that will refute them; almost any given approach can work for some teachers. A wide range of styles should be considered *if* the effectiveness of one's teaching is perceived to be problematic.

One key issue for evaluation arises from consideration of teaching style variables. To label an individual either a "good" or "bad" teacher merely on the basis of the teaching style is to engage in a new form of racism. For example, a person cannot be labeled "good teacher" or "bad teacher" because he or she lectures exclusively. The criterion is learning gain. Improved student growth is the only goal that can justify a style criterion for use in evaluation of staff performance. Such evaluation is costly in terms of money, time, and psychic energy. This does not mean that schools are institutions for learners only. They are institutions for teachers and others. But, the justification of expenditures of energy on staff performance evaluation rests ultimately on improvement of the attainment of the school's goal—that is, student growth in cognitive, affective, and psycho-motor skills, etc.

Some assert that staff performance evaluation is futile; that we cannot know the payoff because it is 18 years down the road. That is another form of faulty reasoning. The best indicator of long-range effects of any given sort is short-range effects. And we ought to use them. It makes no sense to retain a crew that is not getting desired short-range effects in the hope that the group *might* have a long-range effect. So might anyone. It is best to use someone who has more than a mere possibility to his or her name.

The above performance orientation leads some to conclude that the *only* criterion is student achievement. That is not so. Teachers have no basis for treating learners unjustly. Regard-

less of whether or not students grow in the dimensions identified as the goals of the school, a person who perpetrates injustices on students cannot be called a good teacher.

Formative and Summative Evaluation of Staff Performance

The easiest form of performance evaluation to do is *invalid formative* evaluation. Your teaching, or a tape of your teaching, is looked at and the observer tells you what you are doing wrong. This is invalid because it is working from style variables. There aren't any proven style formulas. We need information on the style(s) of teaching for research and improvement purposes. It is valuable descriptive information. But, it cannot be used prescriptively, for summative evaluation.

The next easiest form of personnel performance evaluation is *valid summative* evaluation. It is not too hard to determine the merit of faculty for some personnel decisions. A combination of student ratings, peer ratings of content material, and ratings of the ethicality of the process, collected over extended periods of time, can give a good picture of what a teacher is doing and can assist in systematizing personnel actions. This form is harder than invalid formative evaluation, but not too hard.

Valid summative evaluation done in advance of, and disconnected from, personnel actions regarding assignment and/or retention or dismissal can also serve as valid formative evaluation. This involves the collection of the types of data referred to above. The results are fed back to the teacher, and are discussed by the teacher with a consultant, for possible ideas to be acted upon.

Ideal formative evaluation would consist of evaluation, with diagnostic and therapeutic recommendations. That would be nice if we had the information that would support the therapeutic recommendations. We don't. Some people will complain that it is inhumane to evaluate invalidly and to inflict incompetent teachers on students.

Evaluating for the Purpose of Improving Teaching

There are some things we can do in the interest of helping people improve their teaching. We can go through a list of teaching strategies found to be effective by others. We can look for things that fit in or blend with the style of the person being

evaluated. But we cannot go in and say, "Zap up your teaching. Become zesty, sensitive, and insightful." An applied mathematics professor in an honors school of mathematics was considered a phenomenal teacher by his peers and students. But he mumbled through every one of his presentations. He would not have been highly rated in some other location. We could not go in to him and say, "Become zesty." Our suggestions need to be of the form, "Do you think it would be worth trying to _____?" We have tests that will tell us when we've done well and a repertoire of things that might be tried when we've not. It is strictly a trial-and-error approach for each individual.

There are a number of public schools where it is required that evaluation give teachers *suggestions* for improvement, and that is possible—with no guarantee, of course, that the suggestions will work.

There are two ways to improve a faculty. One is to replace the individuals who are not performing up to standard. The other is to upgrade them. It is not inhumane to say, "My first obligation is to the students." We should not retain or promote someone who is not doing a good job for the students. We are in the business of delivery of services to students. What are our professional obligations?

Our first obligation is to provide professionally well-designed tasks and test items. These should meet both intrinsic (technical) and extrinsic standards of professionality (they should match the course description, for example). And, that's a semi-moral question.

Second, a comprehensive package of materials, texts, labs, field experiences, all built together into a complex that provides as thorough a coverage as you can of the course topics, with the resources that are available, is a requirement.

Third, methods rationale is an important component that we have not previously talked about in any great detail. Given that there is no one best method for all teachers, it does not follow at all that whatever the teacher is doing is okay. It follows that whatever the teacher is doing is not *a priori* identifiable as good. But, of course, each teacher ought to be finding out whether his or her particular methods are any good. A teacher has to get into the business of developing an explanation, a justification of the choice of teaching methods.

This brings us to the fourth point. Self-evaluation is a part of the teacher's professional process obligations.

Fifth, there should be a self-remediation series of suggestions or plans set out by the individual in the light of, and within the rationale for, the results of the self-evaluation—that is, a plan

for self-development. You can see that we are talking about a dossier. We are trying to give the teachers the right to a track record of their teaching activities—something which they own, which presents them as professional teachers, and which is subject to review, and to repudiation. If it says, "I did all the following things," and in fact I didn't, then I not only lose the points for not doing them, but I get black marks for having said I did when I didn't. So, it's not just a fairy tale book; it's not your chance to write a little bit of fiction. It is something that can be used as documentation of your teaching performance.

Sixth, formative evaluation of student work must be done in a clear and thorough way—at least as thorough as time allows. How do we pick that up? By having peers look at the feedback we've written on term papers, to take a college example. Or, we can have another composition instructor look at the feedback on the compositions that were created in a writing course, to take a K-12 case. You can't expect the students to improve their punctuation if you never tell them that their punctuation is no good. It is interesting to note how many teachers insist on getting formative feedback about their teaching performance who do not in turn give feedback to students. The required graduate sociology course at Indiana University was graded on the basis of a single paper. It was normally 50-100 pages long, and it was a required course because it was supposed to get you off and running as a professional sociologist. I remember having a couple of students come to see me once bringing their papers, the 60 pages, and at the end was written, "C+, not very good." And that was it; no other comments through the papers.

Seventh, grading practices must be professionally appropriate. This has nothing to do with the feedback in the sense of detailed feedback, but rather with the overall grade. For example, you cannot keep giving every student an A without special justification. If there are special justifications, they have to be provided. You simply cannot give every student an F without special justification. Why not? Because you're lying to the students. An A means first-rate work—to the students, parents, counselors, and future employers. If I examine the students' papers and find that they are shoddy work, you lie. If you believe you must avoid being a gate-keeper to an elitist institution, don't fool around about it. Make your argument for giving every student an A. We'll listen to the argument. But without a convincing argument for the special case, it's just nonsense when the grades in an institution get to the point where, on a four-point scale, the grade point average is 3.91. In 17,000 grades one year there were 11 Ds and four Fs. Here's the

point where you simply are part of a large-scale con job on future employers. It takes them about five to six years to catch up to the fact that some of these students are A students in nothing whatsoever. I remember that with particular anguish; I once hired someone as a secretary who had an M.A. in psychology from that institution. I found out that she not only couldn't spell, she could not compose sentences and so on.

Eighth, and last on the list of suggestions for professional improvement, is the provision of an *explanation to the students* of past grading practices, to tell them how we measured student growth in previous tests and how it will be measured in tests coming up. Future test performance can be improved by giving examples of good and bad answers with the grades on them.

So those are the considerations of justice and professionality, considerations whose presence you can pick up rather easily from student questionnaires. You need only put in a list of these and say, "Please check if any of these apply." Teachers don't much care for that approach since this is a "black marks list." This is all "downside risk," to use a term from the stock market; you can't get to be a great teacher simply as a result of not being unjust. But, given all that, since the avoidance of injustice or unprofessional behavior is pretty important, one is entitled to check on it—although you are likely to have trouble with the union over a black marks list.

On the professionality topic, there are obviously a number of considerations. For example, very few college teachers can set the items for a good multiple choice test. The people at ETS who are paid very well to do this, and have done it for 20 years, produce about one item an hour in a good working day. People in college do it on an average of one-half hour to an hour per course after five years of practice. When you produce the whole test in an hour, it shows. It's usually a bad instrument—ambiguous, unclear, overcued, and generally tests for rote memory only. That's typical unprofessional conduct. People shouldn't be teaching in college or high school who can't do these minimum things in the professional domain. Those things can be picked up by simply having some professional review of the tests themselves. We should also have peer review—to look at the course content, materials, and tests that are submitted in the dossier.

Finally, among other things, one should inquire: How *enjoyable* was the teaching process? It's very easy to answer. Ask the students; they are the only ones who know. Asking "Did you enjoy it?" is quite separate, however, from "How much did you gain from it?"

Making Comparisons

How meritorious is the teaching process we've identified in the previous lines of inquiry? The key to answering this question is the use of appropriate comparisons. The *most important* appropriate comparison is with other teachers who are dealing with very similar students in the same subject matter. This is easily determined in many cases. For example, elementary school multisections that meet at the same time in the same buildings are an appropriate comparison. But in many other cases, the comparison is less informative because the groups are significantly different. The whole comparison process should be done regularly, after a prior prediction has been made as to which direction the difference will be and for what reasons. You then discount the results to the extent they were predicted on the basis of knowledge of the difference.

The general principle of interpretation in all of this comparison with other teachers is that no personal action is appropriate short of a *repeated* discrepancy of considerably more than one standard deviation above or below the mean, assuming we get a roughly normal distribution of the learning gains. This will often mean that there is no action justified at all because teachers will fluctuate about the mean and fall within this range most of the time. Teachers must be made aware that the mean will drift up when a well-developed program is operating and, hence, that a constant level of performance may not be equated with good performance.

The next most important comparison is with the performance of students taking the previous offering of this class, assuming it dealt with the same subject matter. This is the key *developmental* comparison. It's not of very much use in personnel decisions, but it is very useful for professional development. (You vary the extent to which you go for small group work and so on.) It's a very important comparison in any systematic development process.

In certain cases an important comparison for both formative and summative purposes is with a nonhuman alternative. It may be an audio-visual package, a television course, a CAI presentation, or just simply the text version of the course with some supplementary materials. In other cases a comparison with the teaching performance of the paraprofessional, such as a graduate TA or parent, may illuminate the cost effectiveness of the professional teacher. It may therefore be very important for personnel decisions.

In all the cases I have mentioned above, one should use direct measures of pupil growth whenever possible. But in their absence, student ratings of the amount and quality and so on of what they learned will be the next best substitute.

Integrating and Weighting the Data

Now, given all the data from all the questions and subquestions up to this point, how does one integrate? Some of these efforts will produce more important data than others. Some of them are more reliably obtained than others. There are really just two things to say about this. First, you must use a precise rule. You cannot do a wholistic, interpretive synthesis. If you do, every form of bias known to homo sapiens comes in and with varying degrees; it's quite hopeless. So, use a rule; weight each type of data. Now, which weights? It turns out that the results of weighing them equally are about the same as weighing them two to one, three to one, and so on, essentially because performances on these scales tend to correlate. It may seem attractive to give them differential weights because various stakeholders in this evaluation would like to see some of them given different weights, but it's hard to get agreement. So let the community come up with a recommendation as to how to weight them, with supporting argument, and in general the administration will be able to buy that weighting unless there is some gross blunder involved, like weighting one of them 60 times another.

You may think it's odd to say that using a strict rule is the important thing and the weights are less important. An analogous illustration shows just how critical that point is. In a particular university where the key dimensions for evaluating faculty are research, teaching, and service, data are required on each. Now suppose we don't have any rule for combining the data on research, teaching, and service. Suppose also that the department chair is a keen researcher. When he or she looks at this pile of three kinds of data, with no combining rule, he or she weighs the research results very heavily and the others negligibly. So all this other stuff you were required to submit is, in fact, just ignored, and quite legally. Any weighting rule that didn't have zero coefficients would be better than that.

There is a second problem—contamination of ratings. Even if your teaching is given significant weight in the example above, it is still possible for someone to reinterpret the teaching merit in light of your research performance. An evaluator could look

at your research and not like the looks of it and say, "Intellectually flawed." By association—since the same brain is involved—the *teaching* is intellectually flawed. Your teaching is garbage and doesn't count. That is a simple example of the halo effect, backlash from the evaluator's judgment about your research. If we can't control the extent of that contamination, we will get totally idiosyncratic results. Exactly the same applies to the problem of weighting the various dimensions of teaching effort.

Using Student Ratings

In the postsecondary area, many faculties have moved strongly for a student rating approach, which is, of course, an approximation of student performance. But in other cases it will be a major hassle to get this kind of approach implemented. Not until it becomes standard operational procedure on a campus can it be used to proceed against incompetent teaching. It can be used immediately, however, for *faculty development* by explaining it to volunteers and developing it with local emphasis from that point onward. Not only is it an interesting experiment to perform on one's self, it is also an experiment that documents one's teaching achievements better than any alternative and makes a better dossier for such personnel decisions as promotion and retention. This fact will not be lost on the teacher who is worried about the future fiscal crunches or looking for upward mobility; so volunteers are often numerous.

After the pilot phase with volunteers, you should be able to show that it is an operable system; you then try to get it adopted as a general policy. After that you can proceed against people who shouldn't be there, and reward those who are outstanding. Hopefully, there will be few, if any, in the first category. In general, the advantage of this kind of approach for the teacher is the virtual elimination of idiosyncratic judgment that is often contaminated by personal bias.

This approach can often lead to identification of areas or types of professional performance that need improvement in the self-development effort. The first disadvantage is that this, like any change, involves some disruption and will cost time. Second, any change from a system in which one is doing well involves downside risks. And third, a connected point, this particular system may uncover a weakness hidden by the present one.

For other stakeholders besides the teachers, some of the relevant considerations that affect the political viability of this approach are the following. Parents and members of the general community should be involved in reacting to the proposed system and providing feedback to implement it—especially regarding instances of injustice, but sometimes regarding considerations of content as well. For example, they may help in identifying the usefulness or lack of it of some content choices because of their experiences with the market place. They may also be helpful in reporting on out-of-class contributions to teaching as well as other dimensions of relevant performance such as counseling. Finally, they can help by observing, though (ideally) not voting in, the decision-making procedure itself—i.e., the personnel decision-making procedure. On the whole, the new system should be attractive to teachers, parents, and the community. But where they already have a great deal of power, their participation may cost something. In school districts located on Native American reservations, for example, it sometimes involves a reduction of power.

Students are much better off under this system. They should also have a representative at the decision-making meeting to make sure administrators have followed the rules of the game as promised. The representative, by the way, does not have to be a student but must be chosen by and report back to the students. One can expect to get significant morale changes, especially in unruly high schools, from this commitment to the importance of student input.

How rigid or flexible is this approach? The moment we discover teaching styles that always pay off—for example, in teaching certain new populations being served by the school—we can add them to the repertoire of suggestions for the staff development phase. We can't use them for evaluation; we can't use them for evaluation even if—per impossible—those correlations are 100% correlations because we can't guarantee that they are still at the 100% level. Correlations come from dead data; it's a live person we have to evaluate. But as they get up near the .50 level, we can put those styles at the top of our suggestions list by all means. Just don't penalize anyone for not adopting the suggestion. The minute you do that's de facto proof that you have it in your head as a criterion of merit. You are not allowed to do that, because the evidence to support that conclusion is insufficient. Personnel actions—at least those that are justly rendered—cannot be based on what happened with *other* people. The relevant question is: What happened in *this* teacher's case?

The Implementation Problem

Everyone in the evaluation game has a slice of the implementation apple. The teachers are obviously feeding a dossier, developing a rationale for their methods choice, and so on. They have a large and important input to the item pool. In the best systems, teachers will be doing the grading, knowing that it will be checked by someone else. They are getting the benefit of doing it and, at the same time, learning from not doing it right because someone picked them up on it.

That oversight process also takes the pressure off—it avoids somebody having total responsibility for doing it. Consider a case where you've got an assistant principal or dean in charge of personnel. His or her job calls for going around and visiting the classrooms—the usual approach. The new approach, in the long run, cuts back heavily on the time they spend in classrooms, reducing it to the absolute legal minimum. That will give them some spare time even if, in the first year, they double their time investment on the rest of the process. In later years the process should not take much more than two to three hours per teacher to handle the whole thing.

In most cases you're talking about a few problems in the first year, and you take that out of time you have allocated to staff development, for this *is* part of staff development. Setting up comprehensible standards of merit is the most important part of staff development.

Let's take one situation where the approach suggested here *saves* a lot of time. The average student rating form now in use has, say 20 questions. Many of them have 60; very few of them have less than 20. We'll bring that average down to three, four, maybe five. So all of that time—distribution, filling out, and synthesis—is greatly simplified. The ratings asked for in many of those items are not legitimate for personnel evaluation; they shouldn't be there.

There are other pay-offs. The time spent on developing the skills for observation of teaching will be reduced. The time spent with the teachers in workshops to improve question-asking techniques, and so on, will be reduced. By and large they are replaced by teaching them how to do reasonably decent evaluations of student work as well as other professional tasks. You've got a certain amount of time to play with that comes from the fact that the former system was spinning its wheels. Initially, of course, you will have to schedule a little more time for the changeover, but the process will settle in rather quickly.

You should also think of implementing these changes with a small starting group. Otherwise, it's just too much of a job. Also, it is going to be invalid. You never get these things right the first crack at them. Allow yourself three years to phase it in.

Follow-ups

Follow-ups done over a longer period have both strengths and weaknesses. If we were to do an interview with a 10-year alumnus, much of what we get will be irrelevant. We always face tension when we go to people who have been out of school for awhile, because they are further into life where their schooling pays off, and they supposedly have a better perspective on certain facts.

First, the teachers who taught our hypothetical alumnus probably aren't around anymore so that what the follow-up study has to say can't be brought back to evaluate specific teachers.

Second, the entire program—e.g., course content, methods approach by individual teachers—has probably changed so much, the value of feedback goes down significantly.

Third, so many variables are introduced within a 10-year period, we can be sure that our subject's judgment about causality has been influenced by variables for which we cannot account.

I think you are well advised to downplay the super long-term follow-up studies and concentrate on alumni who are in their first and second years of their vocations. They are very conscious at that stage of what they wish they had gotten, and they are thankful for what they did get that was valuable. Furthermore, their feedback is more likely to be relevant to what your present staff is doing in school.

As with the past, you should be very judicious about explorations into the future. For every year you go into the future, you get a smaller response. As a consequence, you are less and less able to generalize about the present quality of your teaching.

The real payoff from good teaching may be years away, but personnel decisions are here today. They must be made as fair as possible, as fast as reasonable. Students cannot wait 10 years for us to select a faculty. We must select teachers and promote them on the best evidence we can get in the short run. And, it's good enough to improve present faculty evaluation by several orders of merit.

Teacher Evaluation: The Myth and Realities

Carolyn J. Wood and Paul A. Pohland

When asked to identify the utility of summative teacher evaluations, administrators most frequently respond by saying, "to help teachers improve their teaching performance" (Educational Research Service, 1978, p. 179). In the succeeding pages we suggest that this response is, at best, part of the formal doctrine of schools, administrative rhetoric, or, as the title of this article indicates, one of the cherished myths of public education. Our content analysis of teacher evaluation instruments used in the state of New Mexico, both with respect to content items and procedural characteristics, presented a very different reality. In brief, the instruments we examined communicated major concern for organizational maintenance with but minor concern for teaching processes. It is not totally clear why there is such a disparity between the myth and reality, but we speculate that the organizational context in which the principal functions as teacher-evaluator is a major contributor to this phenomenon.

Before proceeding with the analysis and discussion, we need to make explicit two limitations of the study. First, the study focuses exclusively on the instruments used to assess teacher performance. We know very little about how the instruments are used, whether they are the only bases of assessment, and what use is made of the information written on them. In like manner, we know nothing about the skills, preparation, values, or attitudes of the evaluators. Second, the analysis is exclusively quantitative and descriptive. We have no purchase on such qualitative indicators as the value of priority a school administrator may assign to any given analytic category. While it is reasonable to infer, as we do, that the number of items in a given category may correspond well with its perceived importance, this in some ways is a recognized leap of faith.

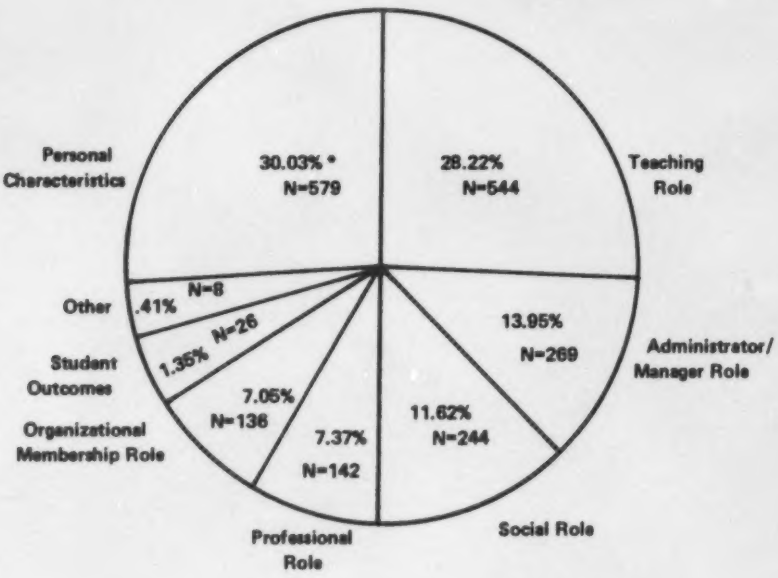
Content Characteristics of Rating Scale Instruments

As in most states, formal teacher evaluation in New Mexico relies heavily upon instrumentation. Consequently, our research initially focused on determining the content characteristics of the teacher evaluation instruments in use in the 88

school districts of the state. Seventy-one of the districts (81%) responded to the request for samples of instruments-in-use. Of the 71, 65 were rating scales while the remaining six were open-ended instruments.

Figure 1

Distribution by Category of Items in Rating Scale Instruments



*Percentages based on N = 1,928

Procedurally, we first scanned a number of the most complex rating scale instruments before initiating a more systematic content analysis of all of the instruments (Holsti, 1968). Scanning revealed that the items were more heavily weighted toward assessing *teachers* in their multiple roles rather than the many aspects of *teaching*. Subsequently we devised a tentative category scheme based on roles and role performance.[1] Its adequacy was tested against another sample of 10 instruments chosen from among the 71. Considerable refinement occurred in the process. Ultimately we re-analyzed all of the data on the basis of the final coding scheme that emerged. The results of the analysis of the 65 rating scale instruments are displayed in Figure 1.

If in reality evaluation were construed as the improvement of teaching performance, one would predict that the teacher role category would account for the majority of items. However, we found that the items in the teaching role category (e.g., "adequate preparation of lesson plans;" "used varied techniques in evaluation of student progress;" "challenges students to think, inquire, and analyze") accounted for only 28.22% of the items in the instruments. Personality characteristics of the teacher (e.g., "flexibility," "enthusiasm," "responsible," "patient") accounted for a higher percentage of items (30.03%).

In addition to focusing on the teacher as a person rather than on teaching, we observed that the instruments required the evaluator to pass judgment on the teacher's entire work world. Items pertaining to the teacher's *social role* (interpersonal relations with students, staff, and community), *administrator/manager role* (student control, classroom management, and reporting/record-keeping activities), *professional role* (academic and professional qualification, attitudes, and development), and *organizational membership role* (conformity to organizational rules and normative expectations) accounted for almost 40% of the items on the instruments. Collectively these categories/items exhibited a clear maintenance rather than teaching thrust.

Within the personal characteristic categories, six subcategories were isolated: 1) personal traits, 2) voice/speech, 3) grooming, 4) physical condition, 5) general characteristics/attitudes, 6) other. The organizational maintenance emphasis of the rating scale instruments became even more pronounced

---

[1]Two categories, "Student Outcomes" and "Other," do not fit into the conceptual scheme of teachers' roles or role performance. However, these categories account for only 1.76% of the content items.

upon examining one of these six subcategories.[2] A more fine-grained analysis of the subcategory "personal traits" yielded four subgroupings: 1) traits related to membership in the organization, 2) traits related to work with students, 3) traits related to the task of teaching, and 4) traits related to the teacher as a person. Table 1 displays the 398 items in the evaluation instruments referring to these groupings.

Data from Table 1 indicate that those traits associated with organizational membership (Grouping 1) combined with those related to the teacher as a person (Grouping 4) comprised the majority (60.05%) of the personal traits assessed. Those items associated with the teacher's instructional role, i.e., work with students (Grouping 2) and the task of teaching (Grouping 3), constituted 39.95% of the items included under personal traits. More particularly, there were three times as many references to punctuality (N = 68) as there were references to interest in students (N = 22).

Further inspection of the personal traits on which teachers are evaluated indicated that items concerned with conservatism and stability were more frequently included by school districts than items oriented toward change. Traits such as punctuality, dependability, responsibility, adaptability, cooperation, accuracy, and loyalty (N = 121) were assessed much more frequently than traits such as leadership, flexibility, initiative, resourcefulness, creativity, originality, and innovativeness (N = 32).

In making the above observations, we are *not* simply suggesting that more items should focus on the teacher's in-class teaching role and fewer on what we have identified as broader organizational maintenance functions. Indeed, the data gathered on current teacher evaluation instruments may be highly pertinent to maintaining the school as an organization and may be useful in making personnel decisions. Any well-functioning organization would seemingly be staffed by individuals who are punctual and loyal, who are able to maintain order and function efficiently, who get along well with others,

---

[2]The subcategory "personal traits" contained 398 items and thus represented 68.74% of the personal characteristics category and 20.64% of the total number of items in the rating scale instruments. Items included under the five other subcategories accounted for the remaining 31.26% of the personal characteristics category: 1) voice/speech (12.61%), 2) grooming (8.98%), 3) physical condition (6.39%), 4) general characteristics/attitudes (2.76%), and 5) the residual category "other" (.52%).

Table 1

Frequency of the Inclusion of Specific Personal Traits Related to Organizational Membership, Students, the Task of Teaching, and the Teacher as a Person

Personal Traits	Number of Times Reported
1. Traits Related to Organizational Membership	
Punctual/prompt	68
Cooperative	12
Dependable	12
Accurate	11
Responsible	10
Loyal	4
Adaptable	4
Flexible	2
Total	123
2. Traits Related to Work with Students	
Interested in	22
Fair	17
Patient	11
Compassionate	10
Positive Attitude Toward	7
Impartial	6
Considerate	3
Tolerant	2
Sympathetic	2
Constructive	1
Democratic	1
Empathetic	1
Humane	1
Courteous	1
Total	85
3. Traits Related to Task of Teaching	
Enthusiastic	43
Initiative	15
Resourceful	6
Leadership	4
Creative	2
Original	2
Innovative	1
Dedicated	1
Total	74
4. Traits Related to the Teacher as a Person	
Judgment	29
Sense of Humor	16
Self-controlled	11
Tactful	10
Poised	10
Sincere	7
Emotionally Stable	5
Honest	4
Exemplar	3
Calm	3
Self-confident	3
Mature	2
Friendly	2
Discreet	2
Other*	9
Total	116

*Each of the following personal traits was reported by one school district: alert, perseverant, patriotic, open-minded, reasonable, optimistic, faces problems honestly, good mental health, positive self-concept.

exhibit professional behaviors, and comply with the rules of the organization. Our point is that such maintenance-emphasizing data are only tangentially related to the primary role of classroom teaching. As such, they appear to be largely irrelevant to the manifest purpose of improving teaching performance and hence provide little leverage for affecting change in teaching practices.

Procedural Aspects of the Evaluation Process

Supporting evidence for the primacy of the organizational maintenance function in teacher evaluation was found when we combined the open-ended and rating scale instruments (N = 71) to analyze the procedural requirements and structural properties of the process. First, only 11 of the 71 districts (15.49%) required the evaluation to be based on more than one classroom visit. Second, only one-third of the documents (23 of 71 districts; 32.39%) indicated specific conferencing requirements. The more typical pattern was the simple requirement that both the teacher and the evaluator sign the evaluation document. Furthermore, only one district required conferencing with tenured teachers following a classroom observation. Third, nontenured teachers appeared to be the primary target of observation.[3] Finally, of the 58 rating scale instruments that used evaluative designations such as "needs to improve," "satisfactory," and so forth, less than half (24 or 41.38%) included designations more positive than "satisfactory," "acceptable," or "average."

If indeed the purpose of teacher evaluation is to improve teaching practices, then it does not seem unreasonable that multiple classroom observations by the evaluator should be required; that conferences should be routinely scheduled with all teachers observed; that both tenured and nontenured teachers should be involved regularly in the process; and that scaling should be broad range and indicative of a concern for excellence. Such characteristics of the evaluation process were not in evidence.

---

[3]Despite some research evidence that beginning teachers are less receptive to administrative feedback than their more experienced counterparts, the practice of specifying neophytes as the prime target of evaluation persists (Tuckman and Oliver, 1968).

Teacher Evaluation in an Organizational Context

Since teacher evaluation occurs within an organizational context, factors affecting the organization as a whole must be taken into account when determining the actual nature of the evaluation process. The number and range of potential contextual factors is considerable. Our discussion will be confined to those that affect the evaluator/principal and correspondingly contribute to the organizational maintenance thrust of teacher evaluation.

By indicating that the manifest purpose of teacher evaluation is to improve teaching performance, school districts communicate an expectation that the principal, as the person normally charged with the responsibility for evaluating teachers, will be instrumental in affecting this "improvement." In a study of the factors currently affecting the principalship, Blood (1978) suggests that conditions present in schools today undermine the principal's capacity to function fully as an "instructional leader." Among the more salient mediating factors, Blood identified the following: increase in the size of schools, increased specialization and professionalization of teachers, dramatic increases in program complexity, diversification of control and bifurcation of the education profession. Additionally, the principal experiences difficulty helping to improve the teaching practices of teachers with more subject matter expertise and experience than he or she possesses. The two major conclusions Blood draws are that 1) the authority and status of the principal have diminished while at the same time the scope of responsibility has increased, and 2) the principal's role has been substantially altered from that of instructional leader to that of manager of a complex organization.

Confirming data on the managerial role of the principal is available from at least two sources. In a companion to the Blood study, Pohland and Higbie (1978) performed a content analysis of the formal job descriptions of principals in New Mexico school districts. The analysis revealed that 85.82% of the items in the job description (N = 1,650) had a managerial focus while only 14.18% referred to instructional leadership. Indeed, the "reporting" requirement alone had a higher frequency of mention than did all the instructional leadership categories and items combined (N = 245 vs. 234; 14.85% vs. 14.18%).

A second source of data is the legal definition of the principalship and its associated set of formal expectations. The New Mexico Public School Code, for example, reads as follows:

Each certified school instructor and administrator in a public school shall:

A. present a valid certificate to the local school board within three (3) months after the beginning of the school year;

B. enforce all laws and regulations applicable to his public school or school district;

C. if instructing, teach the courses of instruction and use the instructional material prescribed for the public school;

D. exercise supervision over students on property belonging to the public school and while the students are under the control of the public school; and

E. furnish reports to his public school officials as may be required.

Additional duties of the school principal are outlined as follows:

In addition to other duties prescribed by law, a public school principal shall be responsible for:

A. assuming administrative responsibility and instructional leadership, under the supervision of the local superintendent of schools, with regard to the discipline of students and the planning, operation, supervision, and evaluation of the educational program of the school to which he is assigned;

B. submitting recommendations to the local superintendent concerning evaluation, promotion, transfer, and dismissal of all personnel assigned to the school to which he is assigned; and

C. performing any other duties assigned him by the local superintendent pursuant to local school board policies.

It is clear from the above that the state places primacy on the managerial aspects of the principalship, the reference to instructional leadership not withstanding. The more general point, however, is that the formal expectations of the principal's job, which are outlined in such documents as the Public School Code and job descriptions, emphasize the managerial role. We would therefore hypothesize that to the extent that administrators are the most influential group in the selection, design, or use of teacher evaluation instruments, such instruments will reflect a basic orientation to organizational maintenance.

Summary

We have suggested here that teacher evaluation serves an organizational maintenance rather than a teaching improvement function. An examination of evaluation instruments currently used indicates this emphasis. First, rather than focusing primarily on the teachers' classroom instructional behaviors, the items on which teachers are evaluated scan the teachers' entire work world. Only a minority of the items are teaching oriented. The items also indicate an organizational preference for stability rather than for change. Second, the procedures used in the evaluation process suggest more concern for assessing "minimal competency" of neophytes than with broad scale instructional improvement. And finally, the formal expectations and conditions which surround the principal/evaluator's role serve to perpetuate the organizational maintenance function of teacher evaluation.

The myth that the purpose of teacher evaluation is to improve teaching may well continue (Guthrie and Willower, 1973). However, we contend that the formal structures embodied in the teacher evaluation process, e.g., the instruments and procedures used plus the requirements and constraints embodied in the principalship, tend to shape the fundamental essence of the process. We are reminded here of Grannis' (1967) statement on the instructional potence of the school structure:

> We are accustomed to thinking of content as the most important learning a school conveys to its students. It is the structure of the school, however, that instructs most systematically, and it is this structure that the students respond to first and remember longest. (p. 15)

In like manner, the various structures surrounding the teacher evaluation process instruct more consistently and more systematically than many of us are willing to admit.

References

Blood, R. E. "Factors Affecting the Principalship Yesterday and Today," in *The Principalship: 1978.* Santa Fe, N. Mex.: Legislative School Study Committee, 1978.

Evaluating Teacher Performance. Arlington, Va.: Educational Research Service, 1978.

Grannis, J. C. "The School as a Model of Society," *Harvard*

Graduate School of Education Association Bulletin, vol. 12, 1967, pp. 15-27.

Guthrie, H. and Willower, D. "The Ceremonial Congratulation: An Analysis of Principals' Observation Reports of Classroom Teaching," *High School Journal*, vol. 56, 1973, pp. 284-290.

Holsti, O. R. "Content Analysis," in G. Lindzey and E. Aronson, eds., *The Handbook of Social Psychology*, vol. II, 2nd edition. Reading, Mass.: Addison-Wesley, 1968.

Pohland, P. A. and Higbie, V. "An Analysis of Job Descriptions of New Mexico Principals," in *The Principalship: 1978*. Santa Fe, N. Mex.: Legislative School Study Committee, 1978.

Tuckman, B. W. and Oliver, W. F. "Effectiveness of Feedback to Teachers as a Function of Source," *Journal of Educational Psychology*, vol. 59, 1968, pp. 297-301.

NSPER: 79

Summary

WILLIAM J. GEPHART
Phi Delta Kappa

ROBERT B. INGLE
University of Wisconsin-Milwaukee

MARY R. CARROLL
Phi Delta Kappa

Wrap-Up: The Basis For Evaluating Teaching

William J. Gephart
Robert B. Ingle
Mary R. Carroll

The presentations and discussions at the three NSPER: 1979 sessions support the contention that planning for the evaluation of teaching is very important. There are many different pieces that must fall into place if the evaluation of teaching is to be carried out in a productive manner. Evaluation always contains an element of threat. To avoid having that threat become debilitating, careful planning needs to take place.

Many people are looking for "the" solution, THE evaluation procedures and policies that will work in almost any location. Based on the discussions in NSPER: 79 we believe that search to be in error. Evaluation is a situation-specific task that rests heavily on, and deals with the values in, a particular location. Transportability of procedures and techniques is extremely difficult. A set of evaluation tools and procedures developed in one location are shaped as they are because of the values held by the people in that setting. To try to use those tools and procedures in another location requires that the same values exist in the second location.

In addition to a unique set of values, there are other unique aspects of the situation that will affect the approach to evaluation, e.g., the specific responsibilities of those to be evaluated, the amount of discretion each exercises over his or her duties and their interpretation, the characteristics of the student body and the community at large, and the nature of the relationships that exist among administrators, professional staff, and other personnel. Each of these singly and in combination will affect the philosophy, procedures, and techniques used in an evaluation.

Even when we create the evaluation system for our own location, we err if we think our design development work is done. Evaluation is a dynamic activity. It will change over time. There needs to be a continuing examination of the evaluation system and the degree to which its goals and objectives are being accomplished.

The participants in NSPER: 79 seemed to accept these three points: 1) the importance of planning, 2) the situation-specific nature of evaluation, and 3) its dynamism. Those three points conditioned the remainder of the discussion of the NSPER: 79 participants. The bulk of the discussion focused on the need for a statement of philosophy about personnel performance evaluation, the need to understand the evaluation process in general, and the process of evaluating teaching specifically. They also expressed some guidelines and principles for evaluating the quality of teaching. Those items will be dealt with in this wrap-up.

Concerns in Planning for the Evaluation of Teaching

Toward a Statement of Philosophy of Evaluating Teaching

The development of a statement of the philosophy of evaluating teaching in a given school is extremely important. It is that statement to which we return when we have questions about procedures and activities to be included in an evaluation plan. A good statement of philosophy should provide us either with the answers to operational questions or with the basis or criteria on which those answers should be developed.

Dale Bolton outlined the content for a statement of philosophy of evaluating teaching. He indicated that it needed to present value bases on which evaluation rests, our assumptions about teaching and about its evaluation, and our expectations about what the evaluation process can and cannot do for us. Readers are encouraged to return to Bolton's statement presented earlier in this monograph for details.

In developing a school system's statement of philosophy regarding the evaluation of teaching, it is probably helpful to identify a set of examples from other places. A number of them can be found. (See for example the statement by Frank Gray about the Hyde Park, New York statement of philosophy presented in *The Evaluation of Administrative Performance: Parameters, Problems, and Practices*, Gephart, Ingle, & Potter, Phi Delta Kappa, Bloomington, Indiana, 1975.) Such statements need to be used to suggest what might be included, not as *the* definitive statement. This point is based on the idea that the values and circumstances in different locations will prescribe differences in philosophy, assumptions, and expectations. We need to be sure that we are building *our* evaluation

plan on *our* value base for *our* situation and not on the base that exists in some other location.

The General Nature of the Evaluation Process

There is confusion about the different processes used for systematic problem solving in education. We are not clear about what those processes do, when they should be used, or what we get when we use them. For that reason it is important that we understand the general nature of evaluation, its components, and its similarities to, and differences from, the research and development processes. We need to sort out confusions among those different processes in order to be more productive in our planning for the evaluation of teaching.

Evaluation is an interesting phenomenon. On the one hand it appears to be very complex, to involve complicated activities and procedures, to be difficult to do, and to be limited to a few who can do it well. Further, it seems that few of us are able to direct others in doing the work well. At the same time, evaluation is an activity we all engage in with apparent ease.

The division between the two sets of circumstances above seems to be the product of several conditions. First is the distinction between evaluation leading to a decision about a personal choice rather than a choice for or by several individuals. Second is the degree to which the decision to be made will have a major impact on the course of the life (lives) involved. Third is the degree of assurance the individual(s) involved have that the evaluative data entering into the decision are objective, reliable, and valid. Fourth is the degree to which the decision can be justified to interested parties. Consider two evaluations: One involves an individual conducting an evaluation preparatory to deciding about whether to move to a distant city to take a new position; the other is an evaluation preparatory to deciding about which dishwasher to buy. Evaluators in these two situations would be faced with considerations that differ substantially in terms of their potential effects and the objectivity, validity, and reliability of the data available to aid in the decision. In the same way a school administrator conducting evaluations prior to making a decision about the purchase of school buses or the evaluation of teachers for retention would be faced with two very different evaluative situations. The data useful in the bus purchase should be relatively exact and reliable. In contrast, the potential for error in the data in a retention decision is much greater.

Sometimes I am involved in decision making in a way in which the decision impacts only on me. At other times I am charged with the responsibility for making a decision in which a large number of people are affected. In the latter case I am making the choice "in trust for" some larger group. If the decision impacts only on me, evaluation seems to be easy. If that impact covers a larger audience (or deals with another's life choice), the activity becomes complex. My desire to make the best possible decision for that group of people (or another individual) makes me want to be sure that the decision was a reasoned choice. It also makes me want to be able to produce the basis for the decision if I am challenged in any way.

Evaluation is one of a number of systematic decision-making processes. It seems to be used best when we confront a problem in which there is a decision to be made, where the decision will impact a number of people, and where the particular choice makers do not know the relative worth or value of the different options on which a choice focuses. As such it is similar to research. Research is a process we use when we want to *know* something in a generalizable or theory sense. When we need an evaluation we want to *know* the relative value of the different options we're considering, a concern that takes place in a specific context, location, and time. We want to know which is the best option at this point in time, given this information. In contrast, through research we want to know the statement that will describe all cases *and* the statement that will be true next year, the year after, and the year after. We recognize that in research we do not attain that goal with great frequency. But nonetheless that is our purpose.

The development process, often confused with evaluation, serves still a different purpose. It is the most efficient strategy to use when our problem is a desire to do some task, a desire that cannot be realized because we lack the tools or procedures to do that task, or those we have will not produce at the qualitative or quantitative level being demanded. Evaluation involves comparison. It involves statements of worth. The development process involves creation, the "engineering" of tools and procedures to do our work assignments.

Real world problem solving is almost always a combination of a need to know in some generalizable sense, the need to choose among competing options, and a need to act. The clear separation of these three processes is seldom found in the real world. The separation made above is for the purpose of clarifying our understanding of them. Readers should not interpret that

statement in a way that reifies the individual processes. They are all involved in the attack and solution of a real world problem.

The Evaluation of Teaching

A contribution of the NSPER: 79 sessions is a structure that can be used to think about the evaluation of teaching. That structure involves the following three dimensions.

1. There is ONE goal and only one goal that can be used to justify the expenditure of large quantities of time, money, and emotional resources in the evaluation of the performance of teachers. That goal is the improvement of student intellectual, social, and maturational development. All of the energy expended in evaluating professional performance is a contribution to the attainment of that goal. If we did not have that concern, it would not be possible to justify the expenditure of time, money, energy, and the anxiety that evaluation creates.

2. There are TWO approaches to the evaluation of teaching that contribute to the attainment of the goal. The first of these is designed solely to facilitate professional growth in the individual teacher. It can be personalized by the statement, "Sometimes we evaluate *my* teaching to give *me* information that *I* will use in *my* decision about how *I* am going to improve *my* teaching." In this case the teacher and some administrative personnel are involved in designing, conducting, and interpreting the results of the evaluation. The teacher is the only person who makes the final decision about what future steps will result from that evaluation.

The second approach to the evaluation of teaching is designed to facilitate administrative decisions about staff assignment and/or retention. This approach can be personalized as follows: "Sometimes we evaluate *my* teaching to provide administrative personnel with information *they* will use in *their* decision about how *they* are going to assign me; whether *they* are going to promote me; or whether *they* are going to dismiss me." In this case the teacher and some administrative personnel are involved in initial discussions about the evaluation to be conducted, but the actual evaluation is designed, conducted, and interpreted by administrative personnel. They are responsible for observing due process procedures and for making the final decision about future actions based on the results of the evaluation.

There are TWO aspects of these TWO approaches to evaluation that are important in different ways for each mode. The

first of these is rapport between the individual being evaluated and those conducting the evaluation. The second is the objectivity, validity, and reliability of the data generated.

As far as possible, open lines of communication in an environment of mutual respect and trust must be fostered. The more positive the rapport between the parties involved, the greater the possibility that a meaningful evaluation will take place. Both parties will be more open to the point of view of the other and thus will perform in a natural and sincere manner. Rapport is very important in the initial stages of evaluation in which the focus is on retention. The teacher being evaluated needs some assurance that those conducting the evaluation respect him or her as an individual. It is also important to those conducting the evaluation. They need a thorough sense of the teacher being evaluated to maintain an appropriate stance in regard to that teacher throughout the remainder of the process. Typically they will have less and less personal communication with the teacher as they move toward a decision.

In the other approach—evaluation for professional growth—rapport between the parties involved is important for the reasons mentioned above and because the contact between the parties remains high throughout the process and is especially so at the interpretive phase of the evaluation. The teacher must have respect for, and some sense of colleagueship with, the administrative personnel involved in this approach to evaluation if the teacher is to make worthwhile use of the evaluative information.

In addition to rapport, the second area of concern in both of these approaches to evaluation is the reliability, validity, and objectivity of the data generated in the evaluation. In evaluation for professional development it is important that the information collected meet these three criteria as far as possible. The teacher who makes adjustments in his or her professional development needs to feel confident that the evaluative data are of high quality. If a teacher is subsequently involved in an evaluation for retention or promotion, all parties involved need as much faith as possible that the evaluation for professional development provided the best possible base for professional growth. In evaluation for promotion or retention, the reliability, validity, and objectivity of the evaluative data collected was critical since the decision being made could have an immediate and significant impact in the life and livelihood of the teacher being evaluated. It is also critical because the administrative personnel conducting the evaluation must be able to justify their decisions to themselves, to the individuals evaluated, and

to the school system in which the evaluations take place. They must be able to demonstrate that they made an appropriate decision either when they decided to dismiss a particular individual or retain another.

Both of these evaluative approaches work toward the improvement of student growth. Both of them exist in all locations. There are times when we try to define one of them, usually the latter one, out of existence. That effort just does not work! Staffing decisions must be made in any institution. They should be carefully weighted, arrived at, and executed as humanistically as possible.

The difficulty seems to spin on the retention/dismissal item. That is where the majority of the anxiety develops. It is estimated that only 2-3% of the people involved in education are, or will become, the evaluatee in a retention/dismissal evaluation effort. Yet, we spend an enormous portion of our evaluation effort on that decision. We need to develop a greater balance.

3. There are THREE categories of variables we need to use in the evaluation of teaching, in either or both of the evaluative approaches. The first class of variables was described by Michael Scriven when he used the term "teaching style" variables. Here we are concerned about lecturing, Socratic questioning, guided learning, discovery learning, mastery learning, etc. As a part of evaluating teaching, we need to know the instructional style and procedures used by the teacher. We need, in dealing with this information, to heed Michael Scriven's admonition that research has not established any particular style as universally best. He said quite bluntly that if you label a person as a good or poor teacher strictly on the basis of the style used, you are engaging in nothing more than another form of racism! Research has not endorsed any specific teaching style. One individual may be very effective as a lecturer; another may not. The second person, however, might be just as effective teaching the same subject matter using a different style. We need to know the kinds of activities used by a teacher if we expect evaluation to help that individual grow as a professional. But, we cannot brand people as good or bad simply on the basis of their teaching style.

The second category of variables was referred to by Macdonald as "employee variables." Here our concern is for variables such as honesty, dependability, ability to get along with others, etc. It was recognized that these items are not teaching per se. They do, however, affect the milieu within

which the teaching takes place. Our concern here is for variables that would apply in any organization.

The third category of variables—context variables—was also introduced in the NSPER: 79 discussion by Macdonald. Because a person works in a school system in South Carolina, there are a number of context constraints on his or her work. These items should be considered in the evaluation of teaching. Included are things such as state laws that prescribe the length of the school year and the school day, laws that tell us what we can and cannot teach, and, in some cases, what materials we can and cannot use. Local policies and procedures also play a role here. The context variables also include the characteristics of the specific group of students with which we work. Again, this is descriptive information that contributes to the evaluation of teaching.

This three-dimensional structure provides the base from which we can design an evaluation plan that will focus on the performance of teachers. We need to be sure that our evaluation touches the three categories of variables that provide information for the two evaluative approaches and that those approaches contribute in our attempts to realize the goal—improvement of student growth.

Some Operational Guidelines

A number of operational principles were discussed by the NSPER participants. Of primary concern was the idea that *most of the time* and *under normal circumstances* evaluation would have as its purpose the provision of information to the individual to be used in self-improvement efforts. In other words the facilitation of professional growth of the individual teacher approach to evaluation is what we are involved with the vast majority of time. That should be acknowledged both in our policy statements and in the procedures we follow. We need to say it *and* do it. Further, our policies and procedures need to make it clear that these two evaluative purposes apply to *all* staff (including administrative, professional, and other support staff).

The second operational point came out quite clearly in discussions with Dale Bolton. He suggested that, if continuing professional growth is not demonstrated over a period of time specified in the school's evaluation policy statement, a written statement of intention to shift evaluative purposes must be presented to the staff member.

Bolton indicated that the individual should be told in writing that a curtain is being drawn on all data generated about that individual's performance prior to the evaluation purpose shift. This is, of course, the desirable ideal. It is possible to disregard records when making a decisio. subsequent to completing the evaluation. It may not be so easy to erase memory. If the same supervisory personnel are involved, some latent aspects of earlier evaluations will inevitably be present in the evaluation to come. Only if a completely different set of evaluative personnel are brought in can the curtain be completely drawn.

The written statement presented to that person should make clear: the nature of the professional growth sought, the resources the school will offer to help the individual achieve the desired growth, the measurement procedures to be used (in terms of how, by whom, and when), the date on which the administrative decision will be made, the options to be considered in that decision, and avenues for redress.

The final point made in relation to procedures for evaluation was an admonition that both process and product approaches for measuring the quality of personnel are needed. If we do not collect data which is descriptive of the product of a staff member's efforts, we do not know if that individual needs assistance or not. If we do not collect process data, we do not know where to focus efforts to try to help the individual improve. Hence, process and product approaches for measuring the quality of professional performance are both needed, but we also need a balance between them.

Summary

Value is the base for evaluation. Evaluation is an activity that involves valuing. And, it occurs within a value complex. As such, procedures for doing it need to be designed to fit within the value complex where the evaluation is to be carried out. Evaluation is often a threatening activity, and it is costly if done well. The only justification for that expenditure is by demonstrating that evaluation contributes to student growth. Evaluation activities get their shape, at least in part, by the purposes they serve and by the values of those involved. In planning for the evaluation of teaching we need to recognize that point by clearly delineating the purposes our evaluation plans are to serve and by direct consideration of the values of the people involved as the plans are developed.

Finally, evaluation is a process that should be evaluated. It is highly unlikely that we will design the perfect evaluation plan for all time. We need to set goals for the evaluation effort and assess the degree to which those goals are attained. The data so generated should be used in decisions about the evaluation procedures and policy.